Methodology

An Unhealthy America:
The Economic Burden of Chronic Disease

Charting a New Course to Save Lives and Increase Productivity and Economic Growth

PART I: The Historical Direct Costs of Chronic Disease

We deploy a cost-of-illness approach to estimate the economic burden associated with the treatment of these chronic diseases. This approach requires information on the types of treatments chosen and their prices by each state. Because treatment options chosen vary greatly by geography, it is necessary to pool actual cost information by category to accurately represent the local economic burden of disease treatment. Because there is no outcome measure, a cost-of-illness approach can be viewed as a low-cost treatment option to bring about the desired outcome.

This approach is chosen as it represents the actual costs of treatment incurred and reflects the asymmetry in treatment options and costs to patients. We compile information by service and product category. Service categories include procedures performed by physicians or other health-care professionals, hospital room, other inpatient care, outpatient care, and nursing home. Product categories include prescription and nonprescription drugs.

Once we have the quantity (population reporting condition) and costs (expenditures per population reporting condition) linked with these treatment choices, we can produce a cost-of-disease treatment for chronic disease. It is important that not only new incidence of disease be included, but ongoing treatment from incidences reported in prior years. The population reporting condition (PRC) should capture this. These would be the historical medical costs, which could be altered in the future through prevention, early detection, and innovation in treatment of chronic disease. We perform this analysis for all fifty states. This modeling system could be extended and implemented for other metros in the future.

Data Sources

We use expenditure information from the Medical Expenditure Panel Survey (MEPS) to perform Stage 1 analysis. MEPS was designed to continually provide policy-makers, health-care administrators, businesses, and others with timely, comprehensive information about health-care use and costs in the U.S. to improve the accuracy of their economic projections. MEPS is unparalleled for the degree of its data and links to specific health-care spending. MEPS collects data on specific health services provided in the U.S., how frequently they are used, their costs, and how they are paid for. Because the data are comparable to those from earlier medical expenditure surveys, it is possible to analyze long-term trends in disease treatment costs.

The Medical Expenditure Panel Survey is a large-scale survey of families, individuals, and their medical providers across the United States. MEPS collects data on each individual's use of medical services and the cost associated with those services.

MEPS data has two major components: a household component (HC) and an insurance component. It also includes a supplemental medical provider component (MPC) and a nursing home component (available only for 1996). The HC is particularly relevant to our analysis because it draws upon a nationally representative sub-sample of households that participated in the prior year's NHIS. Public-use data set in the HC contains demographic characteristics, health conditions, health status, and use of medical services for more than 30,000 persons for each year. Person-based data can be used to make estimates for the civilian non-institutionalized U.S. population by using population-based weighted factors.

MEPS' HC Public use data files consist of consolidated full-year data files and medical event files. A person-level consolidated data file provides expenditure and utilization data for the calendar year from several rounds of data collection. Medical event files provide event-level information for the calendar year on unique household-reported medical events. They consist of seven individual data files characterized by site of service; hospital inpatient stays, emergency room visits, hospital outpatient visits, office-based medical provider visits, home health files, prescribed medicines, dental visits, and other medical expenses. Person-level expenditures associated with a disease type are derived and aggregated from these individual data files.

The Agency for Healthcare Research and Quality (AHRQ) began fielding MEPS in 1996. Since MEPS provides longitudinal information from 1996 to 2003, we can estimate annual costs incurred for all these years at the national level and for each of the four census regions. Lastly, we break out the fifty state expenditures from the four census region figures using the methodology explained below.

For disease information, MEPS data provide both three-digit International Classification of Disease (ICD-9) codes and Clinical Classification Software (CCS) codes. CCS codes were generated by grouping ICD-9 codes into 260 mutually exclusive categories, clinically meaningful disease categories. Most chronic diseases of interest for this analysis are included in these categories: namely, heart conditions, pulmonary conditions (including asthma), hypertension; diabetes, stroke, mental disorders, and cancers.

AHRQ provides useful national and regional-level MEPS summary data tables on expenditure and population reporting conditions for sixty selected chronic conditions from 1996 to 2003. The summary tables are also categorized by individual event files, as denoted by their respective sites of service: outpatient and office-based medical provider visits, hospital inpatient stays, emergence room visits, prescribed medicines, dental expenses, and home heath. Six of the chronic diseases—heart conditions; pulmonary conditions (principally asthma); hypertension; diabetes; stroke, and mental disorders—are presented in this format and may be used for benchmarking. Dental expenses are excluded due to the ambiguity of direct relevance to chronic conditions for this research. The table below provides an example of an expenditure summary table in 2003 for the six chronic diseases mentioned.

Methodology

Expenditure by Site of Service

2003, US$ Billions

Chronic Disease	Distribution by Type of Service: Outpatient and Office-Based Medical Provider Visits	Hospital Inpatient Stays	Emergency Room Visits	Prescribed Medicines	Home Health	Total
Heart conditions	12.6	40.4	3.2	7.3	4.3	67.8
Cancer	23.2	20.4	0.2	1.7	2.9*	48.4
Mental disorders	12.8	7.9	0.8	18.8	7.2	47.5
Pulmonary Conditions	9.6	16.5	1.9	16.2	1.8	46.0
Hypertension	8.5	5.1	0.5	19.6	2.9	36.6
Diabetes	6.4	5.3	0.5	12.4	3.8	28.3
Stroke	1.3	12.0	0.6	0.9	1.6	16.4

* Relative standard error equal to or greater than 30%.

Source: Agency for Healthcare Research and Quality. Total expenses for conditions by site of service: United States, 2003. Medical Expenditure Panel Survey Component Data.

Utilizing MEPS individual event data files associated with sites of services,[1] we estimate expenditure and population reporting condition for specific disease categories that are not available in these MEPS summary tables: namely for breast, colon, lung, and prostate cancer. In order to maintain consistency with the other six diseases, we exclude dental expenditures and other medical expenses that are not directly relevant to chronic disease related to medical costs.

Data Adjustment

A problem with MEPS data is that although MEPS collects data from a nationally representative sample, it is not primarily designed to facilitate smaller geographic-level estimation, such as at the state-level. Only census region (Northeast, Midwest, South, and West) identifiers are available in the HC data files.[2] When we use census region-level data, however, we have to deal with high standard errors due to a sample size bias. For example, the sub-sample size of cancer extracted from the hospital inpatient data file is only 158 in 2003. When broken down into specific cancer types, smaller sample size is likely to yield high standard errors. This bias is more significant at smaller geographic levels.

Hence, to reduce high standard errors, adjustment for outliers is necessary. In particular, the adjustment is applied on the four types of cancer: breast, colon, lung, and prostate. We use MEPS summary tables as a benchmark for this adjustment. Aggregate data on expenditure and population reporting condition for cancer are used as a benchmark when adjusting four specific types of cancer.

[1]. Home health-care expenditures are an exception, since this data file does not provide disease information. Therefore, we use the medical condition file to identify specific disease categories within the file for disease-specific home health care costs.

[2]. Although MEPS provides census region identifiers, there are some data with “inapplicable census regions.” When we estimate specific types of disease, such as lung cancer, the portion of data not assigned to a census region is far too big to be ignored. For example, lung cancer expenditure not assigned to a specific census region in 1999 was 46 percent of national lung cancer expenditure. To correct for this, we re-allocate the “inapplicable” portion back to the census region, based on overall cancer expenditure (or patients) shares by census region.

Methodology

We simultaneously adjust expenditure and population reporting condition across regions. We identify outliers by looking at the share of a specific cancer relative to the total expenditure and population reporting condition for all types of cancer in that region. If the share difference is larger than 10 percent for expenditures and 5 percent for population reporting condition, the data point is adjusted but not excluded. These specific criteria are set to minimize the magnitude of outliers yet maintain attributes from the original MEPS data on expenditure and population reporting condition. The revised regional expenditure and population reporting condition is be scaled up or down so that the total adds up to the national level.

Another problem using MEPS data is that of historical variation across years. In order to obtain representative historical trends, we also have to deal with a variety of factors across years. The table below exhibits historical variations of expenditure and population reporting condition for cancer. As seen in the table, medical expenditures are up and down, compared to changes in population reporting condition across years. We develop a process to adjust for time series outliers across years. The above process does a good job of reducing the influence of outliers for any given year, but not for the variations over time. We compare each year's share of expenditures and population reporting condition for a specific cancer to overall cancer with the eight-year average. After outlier observations are identified, they are adjusted in a similar manner. As before, we scale up or down to match with the U.S. total.

U.S. Cancer (Overall) Expenditure and PRC*

1996-2003

Year	Expenditure (US$ Billions)	PRC* (Thousands)
1996	37.7	9,247
1997	45.5	8,727
1998	35.4	8,952
1999	32.1	9,115
2000	38.9	9,273
2001	45.1	10,316
2002	48.4	10,852
2003	48.4	10,996

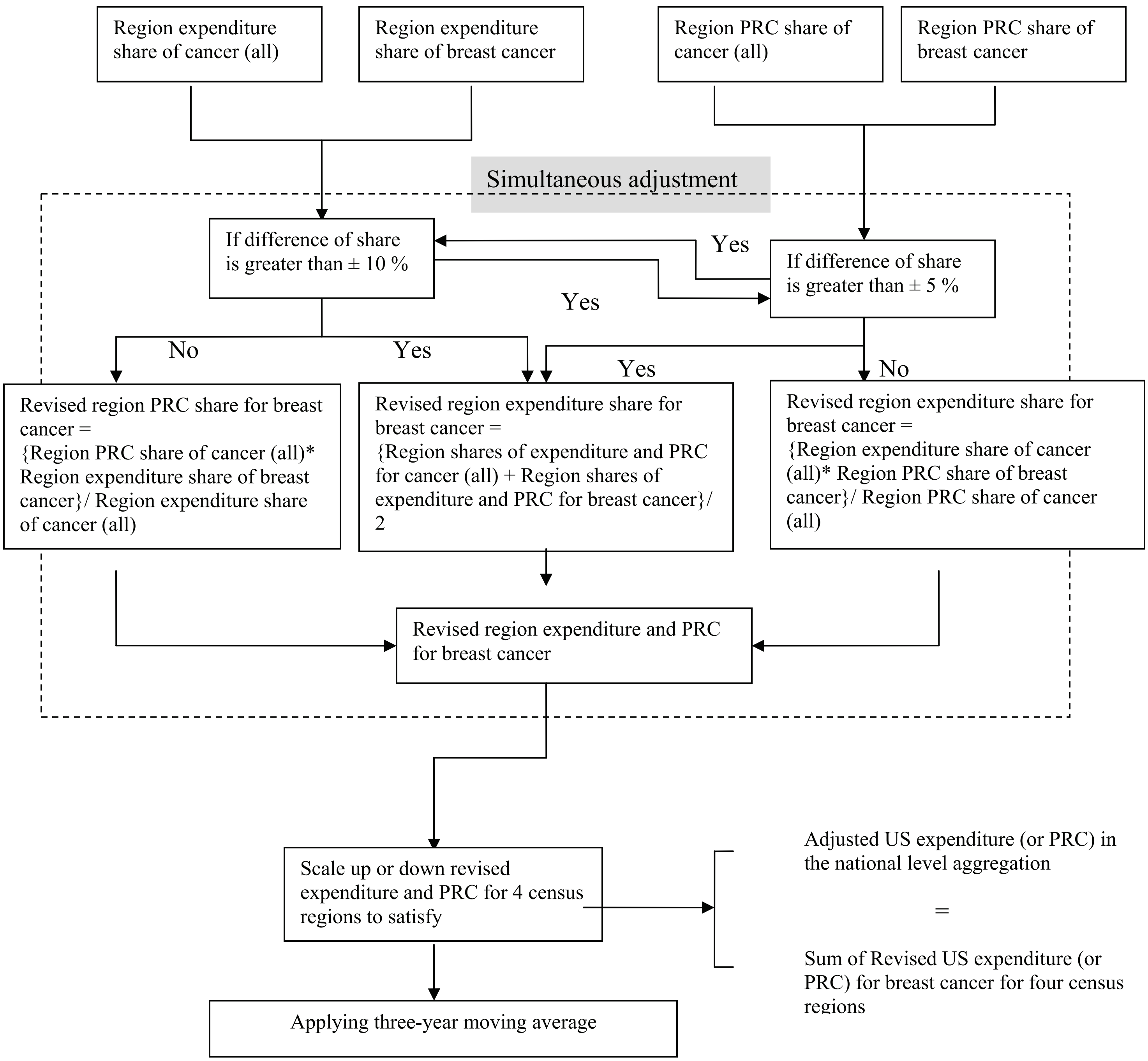
*Population Reporting Conditions

Source: Agency for Healthcare Research and Quality. Total expenses for conditions by site of service: United States, 2003. Medical Expenditure Panel Survey Component Data.

Finally, a three-year moving average is taken for all categories for each region and adjusted to the U.S. total. AHRQ recommends the use of techniques that stabilize trends, such as pooling time periods for comparison (e.g., 1996–1997 versus 1998–1999), working with moving averages, or using modeling techniques with several consecutive years to test the fit of specified patterns over time. Hence, we apply a three-year moving average to smooth variations for both expenditure and population reporting condition. This gives us an estimate of expenditures and population reporting condition by region, along with expenditures per population reporting condition. The following flow chart describes the process. Additionally, the adjustment is applied to other diseases in cases of high variations across years or regions, such as stroke. After we have historically representative disease-specific regional expenditure and population reporting condition, we allocate both to all fifty states, using various state-specific data.

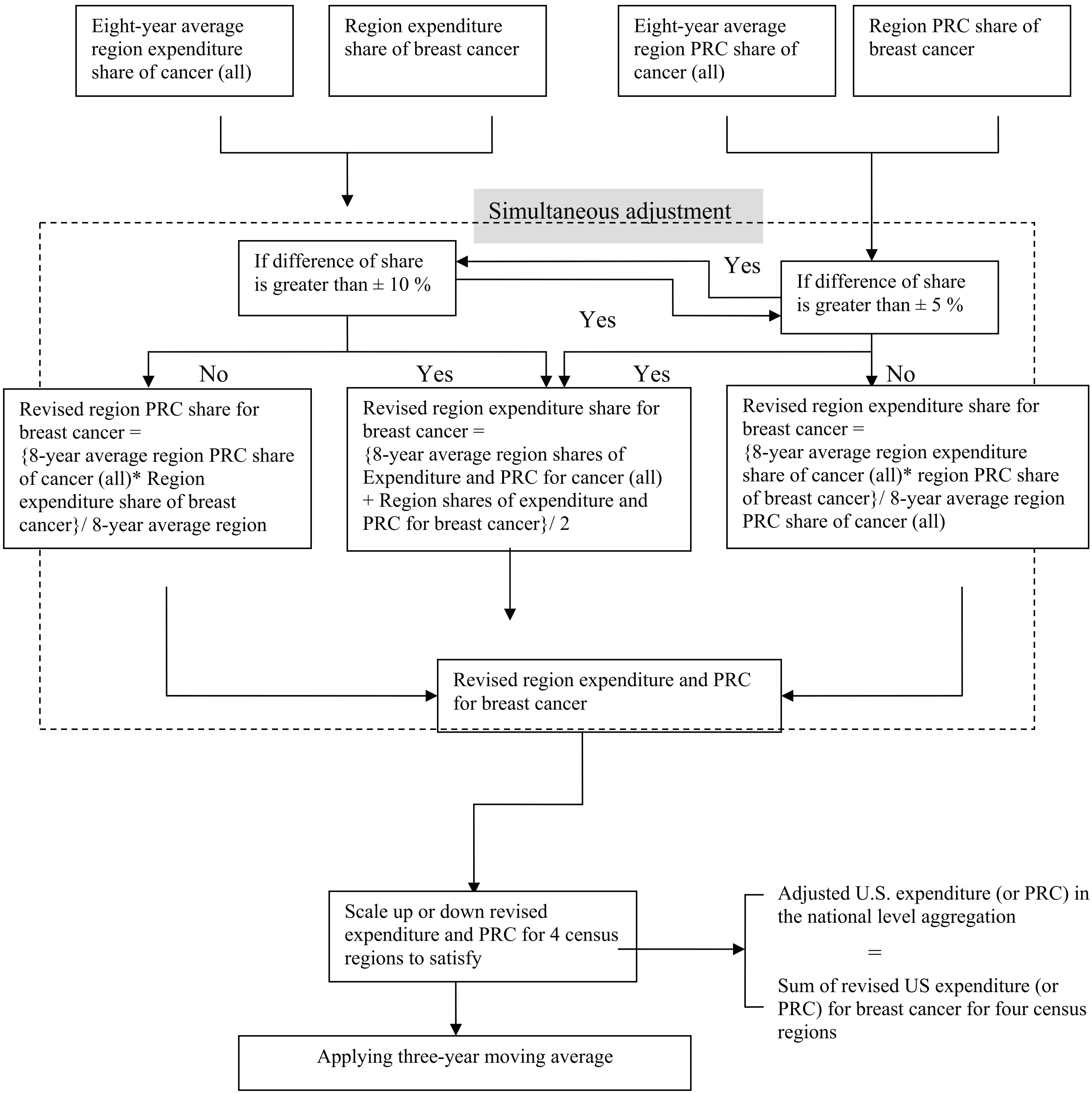
Methodology

Outlier Adjustment Methodology Across Regions: Example of Breast Cancer
***PRC = Population Reporting Condition**



Methodology

Outlier Adjustment Methodology Across Years: Example of Breast Cancer
*PRC = Population Reporting Condition



Methodology

Estimating Disease-specific State Expenditure and PRC

Once we obtain national and census-level representative treatment costs for the eleven disease categories analyzed, the next step is to achieve results at the state level. As previously noted, MEPS provides regional disease-specific treatment costs by site of service, but not at the state level. Meanwhile, the Center for Medicare & Medicaid Services (CMS)[3] does publish personal treatment expenditures at the state level, but only by site of service, not by disease. This data is available from 1980 to 2004.

MEPS data show great variations in expenditures. For example, in 2003, 53.5 percent of MEPS hypertension expenditures (again, derived from "site of service" expenditure tables) went to prescription medications, and just 15.5 percent to hospital care. In contrast, just 10.8 percent of heart disease expenditures went to prescription medications, while 64.2 percent was spent on hospital care.

In order to allocate regional expenditure and PRC to all states, first we created a weighted state per capita expenditure. We applied MEPS disease-specific expenditure shares (by site of services) to state personal health-care costs (by site of service). This produces a "weighted" per capita expenditure by state (weighted by site of service). We next index each state's weighted per capita expenditure against MEPS regional per capita expenditures. Applying this index for each state, we obtained disease-specific expenditure per PRC for all states.

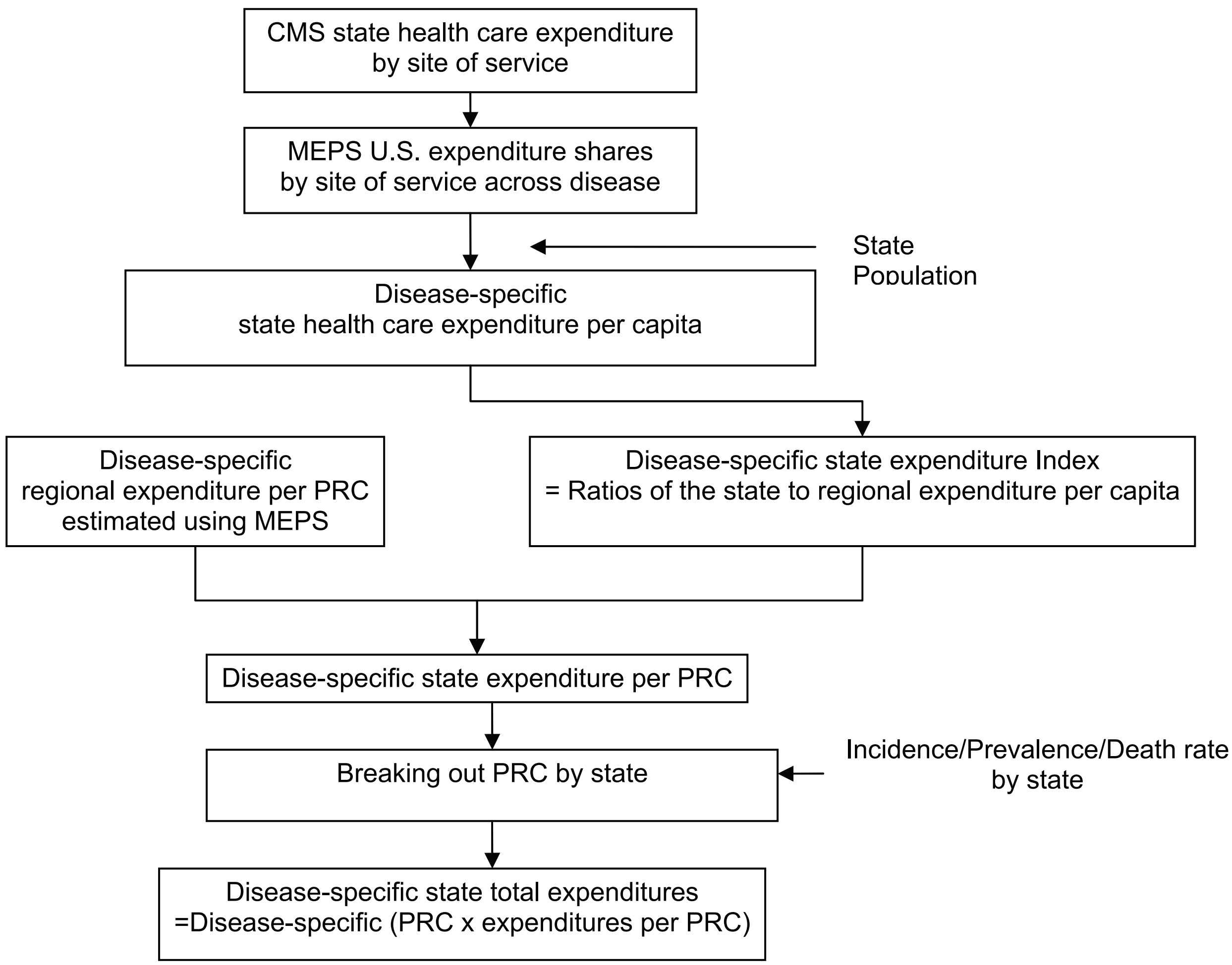
In order to calculate state PRC numbers, we used a combination of incidence/prevalence to break out the number of PRC by state from MEPS regional total. Everything must be benchmarked back to MEPS. Using the incidence/prevalence/death rate, we calculated historical disease-specific numbers of incidence/prevalence/death by state. Then we created state share tables compared to the region. For this step, we used incidence rates for specific cancers; breast, colon, lung, and prostate cancer, and prevalence rates for diabetes, pulmonary conditions, and hypertension. For the rest of diseases, we utilize death rates due to a lack of incidence/prevalence data.

Using disease-specific state shares of incidence/prevalence/death relative to the region, we broke out PRC by state. Then we multiply PRC by expenditures per PRC to derive an estimate of the state total expenditures by disease. The following flow chart attempts to explain this process.

[3]. The Center for Medicare & Medicaid Services is part of the Office of the Actuary, National Health Statistics Group.

Methodology

Estimating Disease-specific State Expenditure
***PRC = Population Reporting Condition**



State Chronic Disease Index

To assess the burden of chronic disease across all states, we create a State Chronic Disease Index. We estimate the PRC per capita and by disease, and then benchmark each state to the state with the lowest PRC per capita. The overall composite is derived by averaging over the benchmark scores for each disease category. That state is assigned a composite value of 100. Thus, a state with a value of 70 means its PRC per capita is 30 percent worse than the top state's.

State-Specific Data Collection

We use state-specific health information to augment the micro-datasets above. We have various state-specific data sources. The Behavioral Risk Factor Surveillance System (BRFSS) was established to overcome the deficiencies of national studies without state-specific information because individual state health agencies have the primary role of targeting resources to reduce behavior risks and their consequent illnesses. BRFSS was initiated with fifteen states in 1984, but all fifty states have participated since 1994. The philosophy behind this survey was to collect data on actual behaviors (rather than on attitudes or knowledge) that would be useful for planning, initiating, supporting, and evaluating health promotion and disease-prevention programs. BRFSS has highlighted key differences in health behavior between states. For example, in 2003, it displayed the wide range between states in the percent of adults who did not exercise, ranging from a low of 15 percent in Minnesota to a high of 30.6 percent in Kentucky. BRFSS is the basic source of incidence or prevalence data.

State cancer profiles from the National Cancer Institute provide historical trends of death rates with demographic characteristics for cancers from 1975 to current. Disease-specific death rates from 1999 to 2002 are also available from the National Center for Health Statistics, Center for Disease Control and Prevention. Meanwhile, new cancer cases from 1997 to 2005 are available from the American Cancer Society and are used as a proxy for incidence rates.

PART II: Projecting Avoidable Direct Costs: Assumptions and Simulations

Despite prevention strategies having a near-term impact on altering unhealthy behavior, the life cycle or cumulative nature of behavior changes and therapies on disease incidence require long-term projections to fully appreciate their potential impact on health-care system cost reductions. We can develop baseline and optimistic alternative scenarios of chronic disease treatment costs based on assumptions from the first-stage twenty years into the future. These projections are based on different PRC and expenditure-per-PRC assumptions. We develop these projections for all fifty states. The avoidable costs are defined as the difference between the baseline and optimistic scenarios.

We develop alternative assumptions on the future path of chronic-disease incidence, prevalence, and PRC, based on best practices in prevention, early detection, and new innovations in disease treatment and management. First, we review current best practices by disease type to determine how more rapid adoption could improve prevention and treatment. The baseline scenario is developed based on a most likely adoption rate. The baseline is formulated based on conservative assumptions regarding new innovations in disease treatment/cures. The optimistic scenario has a more aggressive set of assumptions regarding adoption of current best practices and new treatment/cure innovations. We identify the most likely treatment innovation breakthroughs.

We review the literature on best practices and survey information from different disease groups in prevention/early detection of diseases. Additionally, we review the pipeline of potential treatment interventions.

Model 1: Aging Demographics Only

This section examines aging impacts on chronic diseases. The aging of the baby boomer generation will push demographic factors heavily against reducing overall incidence/prevalence rates over the next twenty years. Changing composition of the population, specifically changes in the number of people or the proportion of people across specific age cohorts, will drive this component of incidence/prevalence forecast. Here we determine how changes in population alone dictate the future trend of incidence/prevalence rates. To derive projections of incidence/prevalence rates for Model 1, we apply the 2003 age-specific incidence/prevalence rates to the census projections from 2004 to 2023.

We used one-year age cohort population projections from the U.S. Census Bureau for the next twenty years. We also used age-specific incidence rates for breast cancer, colon cancer, lung cancer, and prostate cancer from Surveillance Epidemiology and End Results (SEER) of the National Cancer Institute. For other cancers, again, we applied the residuals between all sites of cancers and the four cancers mentioned above. SEER provides incidence rates for the following age cohorts; 0– 49, 50 and over, 55–64, 65 and over, and 75 and over. Incidence rates for the age cohort 50–54 were constructed the given age cohorts of 50 and over, 55–64, and 65 and over.

For other chronic diseases—heart disease, hypertension, pulmonary conditions, diabetes, and stroke—we used age-specific prevalence rates reported by Trends in Health and Aging, Center for Disease Control and Prevention. Unfortunately, due to limitations in obtaining all age cohorts, we examined prevalence projections for the age cohorts of 25– 44, 50–64, 65–74, and age 75 and over. Prevalence specific to the age cohort 45–49 were constructed using the given age cohorts of 44–64 and 50–64.

The following tables include 2003 age-specific incidence and prevalence rates by disease.

Age-Specific Incidence Rates

Per 100,000 Population, 2003

Cancer	Age 0-49	Age 50-54*	Age 55-64	Age 65-74	Age 75 and over
Cancer	94.4	645.0	1035.4	1917.3	2319.0
Breast Cancer	42.4	258.2	319.4	397.9	416.5
Colon Cancer	5.7	60.0	93.1	205.5	339.0
Lung Cancer	4.6	57.1	134.4	325.8	380.7
Prostate Cancer	5.6	184.7	453.9	936.1	834.0
Other Cancers*	60.3	305.6	423.8	743.0	1026.2

* Incidence specific to the age cohort 50-54 were constructed using the given age cohort 50 and over, 55- to 64 and aged 65 and over age cohort.

Source: National Cancer Institute

Age-Specific Prevalence

Percent, 2003

Chronic Disease	Age 25-44	Age 45-49**	Age 50-64	Age 65-74	Age 75 and over
Pulmonary Conditions*	12.6	14.4	17.9	20.7	17.8
Diabetes	2.3	5.9	11.2	18.1	15.8
Hypertension	8.9	19.2	35.1	49.3	54.8
Heart Disease	4.5	7.8	14.6	27.3	36.8
Stroke	0.5	1.1	3.0	7.1	11.6

* Prevalence of pulmonary conditions includes those with asthma, emphysema, and chronic bronchitis.

**Prevalence specific to the age cohort 45-49 were constructed using the given age cohorts of 44-64 and 50-64.

Source: Centers for Disease Control and Prevention

Methodology

Projections of Mental disorders

Due to limitations of age-specific prevalence of mental disorders, we projected 2003 population reporting condition (PRC), which were estimated using data from MEPS. Projections were based on age-specific death rates reported by World Health Organization Statistical Information System (WHOSIS).

Model 2: Pooled Cross-Sectional Model

Historical Behavioral Risk Factors

We obtained most behavioral risk factors relevant to specific types of disease from Behavioral Risk Factor Surveillance System (BRFSS). BRFSS has defined ranges of "at risk" and "not at risk" for each risk factor since 1984. Unfortunately, since historical risk factors for the past twenty years are not directly available on the BRFSS web site, we extracted them using public use data from 1984 to 2003. For consistency, we attempted to use the same question asked in the survey questionnaire where permissible, pertaining to each risk factor within the available data period. Other risk factors (e.g., red meat) are considered.

> *(1) Smoking:* "At risk" for smoking is defined as weighted percentage of respondents who reported that they have smoked at least 100 cigarettes in their lifetime and now smoke. This data is compiled from BRFSS.
>
> *(2) Obesity:* BRFSS does not provide a historically consistent variable for obesity and has changed definitions and weight ranges applied to obesity. Therefore, we calculated historical obesity derived from the body mass index (BMI)[4] in accordance with the most recent definition used in BRFSS. According to the most recent definition of obesity, people whose BMI is equal to or greater than 30 are in ranges of obesity. Therefore, the obesity risk factor is defined as the weighted percentage of respondents specific to those ranges of BMI.
>
> *(3) Exercise:* We used the weighted percentage of respondents who reported participating in any physical activity or exercise, such as running, calisthenics, golf, gardening, or walking for exercise during the past month. The weighted percentage of exercise for both male and female exceeds 65 percent every year. This may be due to individual responses to the question asked and respondents' misinterpretation of the definition of exercise. However, in an attempt to maintain consistency in the question asked, we decided to keep this definition of exercise as a proxy.
>
> *(4) Drinking:* "At risk" for drinking is defined the weighted percentage of chronic drinkers, respondents who reported that an average of two or more drinks per day (or sixty or more alcoholic drinks a month).
>
> *(5) Cholesterol:* We applied BRFSS's most recent definition with respect to the risk factor for cholesterol during the historical period the question was asked. Risk factor for cholesterol is defined as the weighted percentage of respondents who reported that they had their cholesterol checked and were told their blood cholesterol was high by a doctor, nurse, or other health professional. Risk factors for cholesterol are available since 1987.

[4] BMI is computed as weight in kilograms divided by height in meters squared.

(6) *Illicit drug use:* For illicit drug-use, we used the average annual percent of any illicit drug used in past month as reported by United States Department of Human Services.

(7) *Air quality:* With respect to air quality, we used the county-level Air Quality Index (AQI) reported by U.S. Environmental Protection Agency. We aggregated county-level AQI for the each state.

Prevention and Early Screening

(1) Colon cancer screening technology: In tracking the latest and most widely used technology with respect to colon cancer screening strategies since 1988, BRFSS has revised the questionnaire whenever new screening technology has been introduced. BRFSS changed the question of "ever having had a proctoscopy" between 1988 to 1995, to "ever having had a sigmoidoscopy or proctoscopy" between 1996 to 1998, to "ever having had a sigmoidoscopy/colonoscopy" from 1999 to 2000, and finally revised it to "ever having had either sigmoidoscopy/colonoscopy" from 2001 to 2003.

We defined the adoption of colon cancer screening strategies as the weighted percentage of respondents who answered "yes" with respect to the questionnaire for each year. We expect the changes in adoption rates with respect to changes in the questionnaire to reflect the influence of technological innovation in colon cancer screening strategies.

(2) Hypertension drugs: Historically, hypertension drugs have made a significant impact in the treatment of cardiovascular disease by reducing the probability or onset of such conditions. By effectively lowering high blood pressure, the chances of a heart attack can be significantly minimized.

With the first drug introduced in 1952, the number of hypertension drugs currently on the market has increased to fifty-three. In our model, the increasing growth in the number of drugs available, particularly throughout the 1980s, had a tremendous impact on the prevalence of heart disease, stroke, and hypertension.

Hypertension drugs comprise five classes: alpha blockers, beta blockers, calcium channel blockers, ACE inhibitors, and diuretics. Cumulatively, the number of hypertension drugs introduced has increased dramatically since 1952. Relevant information was collected from the Federal and Drug Administration (FDA).

Cross-Sectional Regressions by Disease

Pooled, cross-sectional regressions were used in determining the impact of various demographic and behavioral risk factors on disease-specific incidence/prevalence. In total, eleven cross-sectional regressions were performed, one representing each chronic disease.

The dependent left-hand side (LHS) variables comprise incidence, prevalence, or death rates, depending on the chronic disease. Explanatory right-hand side (RHS) variables include both demographic and behavioral risk factors, such as smoking and obesity.

Methodology

We also incorporated variables such as income and education to control for socioeconomic status by state. Data on state median income and percentage of the higher educated (population with bachelor's degree or higher) were compiled by Bureau of Economic Analysis, U.S. Census Bureau, and Economy.com. We used disease-specific risk factors by state from the BRFSS public use data and demographic factors reported by U.S. Census Bureau.

In formulating disease-specific cross-sectional regression models, we construct three-year pooling state-level data (for all fifty-states) from 2001 to 2003, which provided for greater and more significant variation across risk factors. In performing our empirical analysis, we chose double log specification of regression models. Variables used to capture the impact of prevention and effective treatment are discussed later since they were estimated using national level data outside of our cross-sectional model.

(1) Breast Cancer

The female population 65 and older and the weighted percentage of female obesity significantly explain breast cancer incidence. As expected, the older females and ones with a BMI greater than 30 are likely to have higher incidence of breast cancer. Aging comes up as the most significant risk factor, as seen in the table below.

Dependent variable: Log (Incidence for Breast Cancer)

Sample Period 2001 to 2003

Variable	Coefficient
Constant	-11.4162*** (-12.95)
Log (Median Income)	0.7396*** (79.23)
Log (Female Population Aged 65 and over)	0.9447*** (8.32)
Log (Percent of Obesity for Female)	0.3398*** (3.63)
Sample Size	153
R-Square	0.9801

*** Statistically significant at the 1 percent level or better

(2) Colon Cancer

Smoking represents the most significant risk factor. The population 65 and over is also significant. Obesity and a higher percentage of "at risk" smokers are likely to increase incidence. A 1.0 percent change in smoking prevalence results in a 0.5 percent incidence change in the same direction. Since exercise is significant, we may conclude that incidence decreases with moderate exercise.

Dependent variable: Log (Incidence for Colon Cancer)

Sample Period 2001 to 2003

Variable	Coefficient
Constant	-10.5191*** (-15.18)
Log (Median Income)	0.47908*** (6.7)
Log (Population Aged 65 and over)	0.4330*** (9.33)
Log (Percent of Obesity)	0.0167 (0.18)
Log (Smokers at Risk)	0.5291*** (10.88)
Log (Percent of Doing Exercise)	-0.4531** (-2.86)
Sample Size	153
R-Square	0.9907

*** Statistically significant at the 1 percent level or better

** Statistically significant at the 5 percent level or better

(3) Lung Cancer

Both smoking and the population 65 and over exhibit high significance. A 1.0 percent change in smoking prevalence leads to a roughly 1.0 percent incidence change in the same direction. Lung cancer probability increases with age, reflecting the cumulative effects over a lifetime.

Dependent variable: Log (Incidence rate for Lung Cancer)

Sample Period 2001 to 2003

Variable	Coefficient
Constant	5.7851*** (-7.72)
Log (Median Income)	0.0349 (0.46)
Log (Percent of Population Aged 65 and over)	0.2401*** (3.57)
Log (Percent of Smoker at Risk)	0.9755*** (12.5)
Sample Size	153
R-Square	0.578

*** Statistically significant at the 1 percent level or better

(4) Prostate Cancer

Prostate cancer tends to occur more often in African Americans and men 65 and over. Male obesity is also a significant determinant. A 1.0 percent change in obesity prevalence leads to a 0.5 percent incidence change in the same direction.

Dependent variable: Log (Incidence for Prostate Cancer)

Sample Period 2001 to 2003

Variable	Coefficient
Constant	-9.2430*** (-10.39)
Log (Median Income)	0.5349*** (6.55)
Log (Male Population Aged 65 and over)	0.4034*** (6.16)
Log (Percent of African-American Population)	0.0306** (2.43)
Log (Male Population with Obesity)	0.5204*** (7.93)
Sample Size	153
R-Square	0.977

*** Statistically significant at the 1 percent level or better

** Statistically significant at the 5 percent level or better

(5) Other Cancers

Since "other cancers" are not specific to one type, we test against various behavioral and demographic factors. Obesity, smoking, and cholesterol display high significance in "other cancer" incidence, but demographic factors, particularly aging, also yield high correlation. A 1.0 percent change in obesity prevalence leads to 0.3 percent incidence change in the same direction.

Dependent variable: Log (Incidence for Other Cancer)

Sample Period 2001 to 2003

Variable	Coefficient
Constant	-7.7110*** (-8.16)
Log (Median Income)	0.3689*** (4.04)
Log (Smokers at Risk)	0.2142** (2.16)
Log (People with Drinking at Risk)	0.0282 (0.61)
Log (People with Cholesterol at Risk)	0.1263** (2.24)
Log (People with Obesity)	0.3036*** (5.93)
Log (Population Aged 65 and over)	0.2641*** (3.34)
Log (Percent of Hispanic Population)	0.0357** (2.55)
Log (Percent of African-American Population)	-0.232** (-2.01)
Sample size	108
R-Square	0.9893

*** Statistically significant at the 1 percent level or better

** Statistically significant at the 5 percent level or better

(6) Heart Disease

Due to the lack of state-level prevalence/incidence data, we use death rates as a proxy for the dependent variable. Age and obesity are the most significant factors, followed by smoking. Exercise appears to decrease the risk significantly. A 1.0 percent increase in physical activity prevalence leads to a 1.2 percent decrease in heart disease death rates.

Dependent variable: Log (Death rate Due to Heart Disease)

Variable	Coefficient
Constant	8.1278*** (19.85)
Log (Median Education)	0.3261*** (3.93)
Log (Percent of Population Aged 65 to 74)	0.4954** (2.16)
Log (Percent of Population Aged 75 and over)	0.5046*** (3.88)
Log (Percent of Smokers at Risk)	0.3214*** (3)
Log (Percent of People with Obesity)	0.5243*** (3.72)
Log (Percent of People Doing Exercise)	-1.2436*** (-4.99)
Sample Size	51
R-Square	0.909

*** Statistically significant at the 1 percent level or better

** Statistically significant at the 5 percent level or better

(7) Hypertension

We use prevalence as the dependent variable. Age and obesity seem positively and significantly correlated. Exercise appears to reduce occurrence of hypertension and was found to exhibit a notable and separate impact on hypertension from its associated link to obesity. A 1.0 percent increase in physical activity prevalence leads to a 0.3 percent decrease in hypertension prevalence.

Dependent variable: Log (Prevalence for Hypertension)

Sample Period 2001 to 2003

Variable	Coefficient
Constant	-0.4186 (-0.88)
Log (Median Income)	-0.0244 (-0.53)
Log (Percent of Population Aged 50 and over)	0.3735*** (6.12)
Log (Percent of People with Obesity)	0.3101*** (5.74)
Log (Percent of People Doing Exercise)	-0.9189*** (-7.7)
Sample Size	153
R-Square	0.6804

*** Statistically significant at the 1 percent level or better

(8) Diabetes

The population 65 and over appears to be the most significant factor increasing the prevalence of diabetes, which shows the cumulative impact of more over the life cycle. A 1.0 percent change in population 65 and over leads to almost 0.8 percent prevalence change in the same direction. Among behavioral risk factors, obesity has the strongest relationship with diabetes, apparent from the highly significant coefficient.

Dependent variable: Log (Prevalence for Diabetes)

Sample Period 2001 to 2003

Variable	Coefficient
Constant	3.0481*** (4.85)
Log (Median Income)	-0.3751*** (-5.61)
Log (Percent of Population Aged 50 and over)	0.7520*** (8.26)
Log (Percent of African-American Population)	0.0842*** (10.93)
Log (Percent of People with Obesity)	0.3647*** (4.39)
Sample Size	153
R-Square	0.7472

*** Statistically significant at the 1 percent level or better

(9) Asthma

The onset of asthma[5] typically occurs to individuals under 40. Thus, we do not include age as a variable. Asthma is likely to be more prevalent among the Hispanic population, but the disease impact in that population is not large, as indicated by the small coefficient. Smoking and air quality appear to be major risk factors. A 1.0 percent change in smoking prevalence results in a 0.6 percent asthma prevalence change in the same direction. Air quality also seems to have a fairly significant impact.

We used the cross-sectional regression model for asthma as a proxy of pulmonary conditions (including asthma) in this research.

Dependent variable: Log (Prevalence for Asthma)

Sample Period 2001 to 2003

Variable	Coefficient
Constant	-1.4502*** (-3.96)
Log (Percent of Hispanic Population)	0.0372** (2.11)
Log (Smokers at Risk)	0.5877*** (7.9)
Log (AQI)	0.3252*** (4.82)
Sample Size	150
R-Square	0.9688

*** Statistically significant at the 1 percent level or better
** Statistically significant at the 5 percent level or better

[5]. BRFSS provides data only on asthma, as opposed to pulmonary conditions.

(10) Stroke

We use the death rates as the dependent variable due to the limitation of state prevalence data. Smoking appears to be the most significant behavioral risk factor, as indicated by its highly significant and large coefficient. A 1.0 percent change in the number of smokers results in over a 0.6 percent death rate change in the same direction.

Dependent variable: Log (Death Due to Stroke)

Sample Period 2001 to 2003

Variable	Coefficient
Constant	-3.9948*** (-3.53)
Log (Percent of Population Aged over 65)	0.3238*** (2.97)
Log (White Population)	0.3577*** (6.21)
Log (Median Income)	0.0236 (0.19)
Log (Smokers at Risk)	0.6319*** (10.65)
Log (Percent of Population with Obesity)	0.1009 (0.67)
Sample Size	153
R-Square	0.9714

*** Statistically significant at the 1 percent level or better
** Statistically significant at the 5 percent level or better

(11) Mental disorders

We use death rates as the dependent variable due to limited data on incidence or prevalence rates. Heavy drinking and illicit drug use appear to increase mental disorders. Age is also a significant factor, with statistical significance at around 10 percent. A 1.0 percent change in the population 65 and over leads to over a 0.2 percent change in the death rate.

Dependent variable: Log (Death Rate of Mental Disorders)

Sample Period 2001 to 2003

Variable	Coefficient
Constant	-6.6931*** (-4.18)
Log (Population Aged over 65)	0.2489 (1.89)
Log (Median Income)	-0.0606 (-0.42)
Log (Drug Use)	0.0656 (0.5)
Log (Percent of Drinkers at Risk)	0.1954** (2.15)
Sample Size	153
R-Square	0.134

*** Statistically significant at the 1 percent level or better

** Statistically significant at the 5 percent level or better

Risk Factor Projections

For the next step, we must have risk factor projections for the next twenty years. For demographic factors, we used census projections reported from U.S. Census Bureau.

For behavioral risk factors, first we projected 2003 national-level values, using relevant baseline and optimistic assumptions of the end points for 2023. Then we calculated projections from 2004 to 2023 based on the historical trends. We obtained state risk-factor projections based on trends of national risk factor projections. Finally, we applied state-level risk factor projections of baseline and optimistic scenarios to cross-sectional/national time-series regression models in order to project the future trend.

Assumptions for risk factor projections

Smoking

Baseline assumptions: Smoking declines at the same rate it fell between 1985 and 2005. The percentage of “at risk” smokers (individuals who smoke at least 100 cigarettes over their lifetime and who still smoke) will fall to 19 percent in 2023.[6]

[6]. Smoking statistics come from a BRFSS survey question.

Optimistic assumptions: Smoking declines at the same rate it dropped between 1965 and 2004. In 2023, approximately 15.4 percent of the adult population will smoke.

Obesity

Baseline assumptions: The baseline scenario calls for the rate of obesity to moderate and begin to plateau around 2015. We assume that the prevalence of overweight conditions grows at about half the historical increase, or 43.6 percent, in 2023. Obesity increases to 28.7 percent in 2023.

Optimistic assumptions: A change in unhealthy behaviors, combined with therapeutic-compound effects, will significantly influence the upward trends of obesity. Wellness programs will affect BMI through diet, exercise, leisure activities, and education. Overweight prevalence will drop to 32.2 percent of the population in 2023, and obesity will fall to 19.4 percent. We assume that male and female obesity will follow the same trends. Obesity prevalence will decline to 19.7 for men and 19.2 percent for women in 2023.

Exercise

Baseline assumptions: The percent share of the population engaged in physical activity will increase gradually, from 75.4 in 2003 to 77.9 in 2023.

Optimistic assumptions: The population share engaged in physical activity will increase to 83.3 percent by 2023.

Drinking

Baseline assumptions: The "at risk" percent of the population remains unchanged at t 5.8 percent.

Optimistic assumptions: The percentage of "at risk" population decreases steadily to 4.2 percent. Raising awareness of the adverse effects—in particular, the links to chronic diseases—will lead to lower alcoholic consumption per capita.

Cholesterol

Baseline assumptions: We expect the population share with high cholesterol to stabilize at around 42.2 percent in 2023.

Optimistic assumptions: Increased awareness of diet and nutrition, and their impacts on healthy aging, will help lower cholesterol levels. We assume that the population percentage with high cholesterol will drop to 31.5 in 2023.

Illicit Drug Use

Baseline assumptions: We assume that the usage trend will plateau in the next twenty years, attributable to increased awareness of the adverse effects of illicit drug use and stricter law enforcement policies. The number of arrests as a share of the total population will climb to 0.64 percent in 2023, an increase of 14.2 percent from 2005.

Optimistic assumptions: We assume that the number of arrests as a share of the total population will decline at a faster rate, ultimately reaching 0.57 percent by 2023.

Air Quality

Baseline assumptions: To capture a historical trend, we create a national air quality index that captures growth in fuel demand (as measured in BTUs) and population, based on data from the Environment Protection Agency (EPA). We assume that demands for fuel will increase as the population grows, causing the index to follow its historical trend. As a result, air quality worsens steadily, from 40.1 in 2003 to 58.4 in 2023, an increase of 46 percent.

Optimistic assumptions: We assume a net reduction in air pollution and other allergens and irritants attributed to more environmentally friendly alternatives to fuel and/or incentives, such as ridesharing and low-emission vehicles. Air pollution increases at a slower pace, reaching a level of 53.5 on the index in 2023.

Model 3: The Path of Screening and Treatment Innovation

National Level Regressions

Model 3 builds on Model 2, which calculated assumptions of risk factor trends into the aging demographic projections of Model 1. Now we estimate the positive values of various screening and treatment (therapeutic compounds) innovations. These impacts can be estimated into baseline and optimistic projections of prevalence and incidence.

Because state-level data are limited, we rely on national-level data to build time-series regression models. There is available data for just six of the eleven disease categories under study: colon and prostate cancer, heart disease, hypertension, mental disorders, and stroke.

Time-series-based national-level regressions were calculated in the cases where data on prevention and early screening were available. For example, the number of hypertension drugs came in significant when tested against prevalence of heart disease, stroke, and hypertension. Since these data were available historically only at the national level, these results at the national level were overlaid into results from the cross-sectional model discussed above.

First, we estimated time-series-based regressions at the national level. The model included historical data on behavioral and demographic risk factors (e.g., smoking, obesity, population 65 and over, etc.) and pertinent variables representing prevention and early screening (e.g., hypertension drugs, colonoscopy screenings, etc.).

The next step involved running the estimated coefficients through our projected risk factors and preventative and early screening assumptions. From here, we were able to derive projections on incidence/prevalence at the national level. Finally, annual percent changes were applied to our corresponding results estimated from our cross-sectional model.

Methodology

Projections of screenings and Treatment Innovation

Cardiovascular (hypertension) Drugs
Baseline assumptions: We expect the cumulative number of FDA-approved hypertension drugs on the market to reach 57 by 2023, an increase of 21.3 percent over the twenty-year period. Furthermore, we expect growth in FDA-approved drugs to moderate throughout the baseline projection period, as tremendous strides in therapeutic compounds have already occurred in the past twenty years.

Optimistic assumptions: We expect the cumulative number of FDA-approved hypertension drugs on the market to reach fifty-nine by 2023, an increase of 25.5 percent over the twenty-year period. With respect to drugs available to treat heart disease, the cumulative number would increase sixty-four, an increase of 36.2 between 2003 and 2023. Here we assume wider variety and options to treat such disease.

Colon Cancer Screening Technology
Baseline assumptions: We utilize historical figures in colon cancer screening trends to project the baseline scenario. From 1988 to 2003, screening for colon cancer increased by 17 percent, according to survey respondents. We expect that an increase of 13.7 percent in colon cancer screening in the following fifteen year. By 2023, 60.3 percent of the adult population, or almost two out of every three Americans, will be screened for colon cancer.

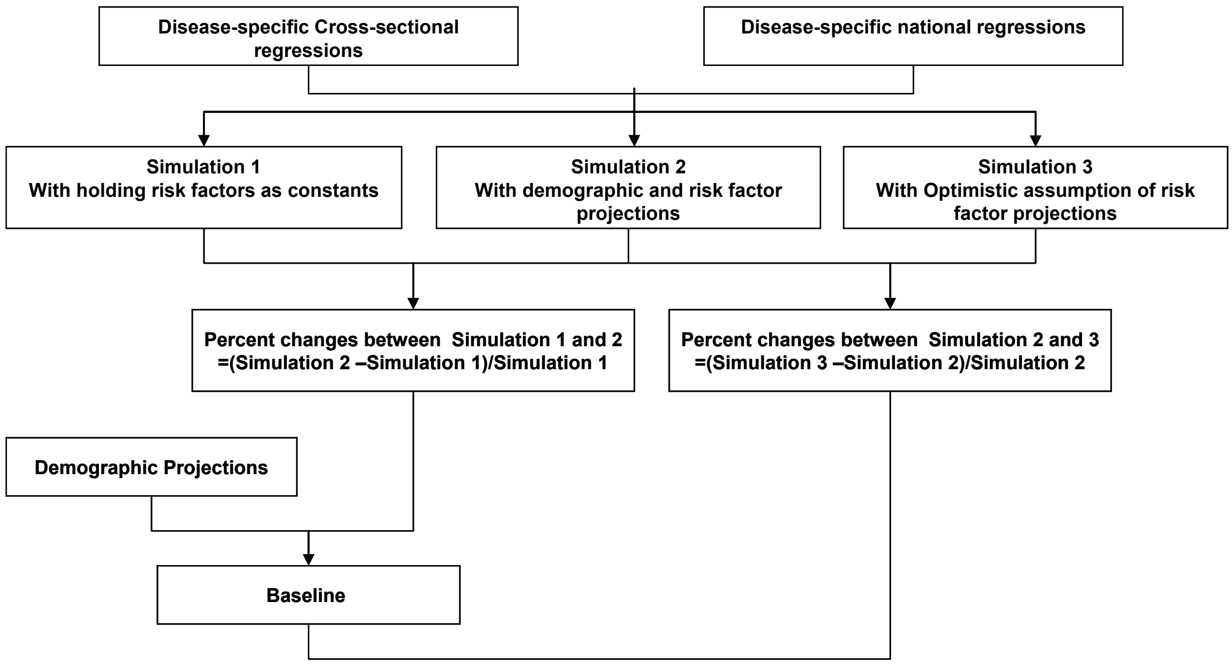
Optimistic assumptions: In the optimistic scenario, we expect more aggressive action in promoting more early screening, especially as more advanced technology becomes available. Here we project the percentage of adults receiving screening to increase to 69 percent.

Simulations Based on the Three Models

Using the preceding models—two of which include baseline and optimistic assumptions—we ran simulations that enable us to build twenty-year projections for overall baseline and optimistic incidence/prevalence.

In order to do so, we constructed through three simulations: 1) simulation 1, accounting for changes in demographic factors only, holding behavioral risk factors at their 2003 values, 2) simulation 2, accounting for baseline projections of demographic risk factors, and screening and treatment innovation, and 3) simulation 3 accounting for projections of demographics and optimistic risk factors, and screening and treatment innovations.

The following diagram helps to illustrate the process.



The first simulation accounts for changes in *demographic factors*[7] (age and race) only, holding behavioral risk factors at their 2003 values. This simulation should be conceptually consistent with demographic driven (aging) projections obtained in Model 1.

The second simulation accounts for *baseline* projections of Model 2 and Model 3, accounting for *behavioral risk factors* plus available screening and treatment options. We apply the percent changes between the results the first and second simulations to the age-driven demographic projections established in Model 1. This will give us final baseline incidence and prevalence projections.

The third simulation is the same as the second but accounts for *optimistic* risk factor projections plus available *optimistic* screening and treatment options. Similarly, optimistic and incidence and prevalence projections are completed by applying the percent changes between the second and third simulations to the final baseline established in the second simulation.

Projections of Population Reporting Condition (PRC)

Finally, applying health-care cost growth on expenditure, we projected 2003 disease-specific expenditure per PRC estimated in Part 1. This enabled us to obtain disease-specific expenditure projections by applying PRC to expenditure per PRC for the period twenty-year period.

[7]. See the table "Pooled Cross-Sectional Models" above.

Methodology

We projected 2003 MEPS Population Reporting Condition based on annual growth rates of incidence/prevalence projections from proceeding baseline and optimistic scenarios. In fact, in the case of cancers, PRC would be greater than new cases because PRC includes people with diagnosis at a certain time, as well as new cases. Moreover, projections of PRC should also consider people who leave the sample size from death or from complete cures, even if those are not many. However, under the constraint that projections of disease-specific death rates and cure rates are not available, it is not feasible to count people who leave from the survey. Therefore, similar to other diseases, we projected cancer PRC for based on annual growth rates of incidence projections.

Finally, we project 2003 state PRC from the regional MEPS conversions, using state variations from the pooled cross-sectional models. Then the sums of disease-specific state PRC are adjusted with U.S. MEPS control totals for each year.

Projections of Health-Care Cost Growth

Once disease incidence and prevalence rates have been converted to PRC, we can examine expenditures by PRC and total disease expenditures.

We assume that health-care cost growth will follow projections of the Centers for Medicare and Medicaid Services (CMS). Health-care cost growth for the "optimistic" scenario is 0.5 percentage point lower than that used in the baseline projections.

To make disease-specific expenditure projections, we adjust the projected inflation rates to account for future costs associated with four specific sites of service (again, we use 2003 MEPS data). The four sites of service include (1) outpatient and office-based visits; (2) home health care; (3) prescription drugs; and (4) hospital inpatient visits, including emergency room services.

Assumptions for Expenditures per PRC in the Baseline Scenario

(1) Health-care cost growth by site of service: The CMS projects a "personal health care" price deflator,[8] which is its overall rate of inflation for the private health sector. The CMS does not report health-care cost growth by "site-of-service." To estimate health-care cost growth for our four categories, consistent with the CMS projection of overall health-care inflation, we extract historical data and projections for specific health-care price indexes from Global Insight.[9]

(2) Sectoral forces: In "Health Spending Projections Through 2015: Changes on the Horizon,"[10] Borger and her colleagues describe the factors they considered when CMS assembled its projections.[11] On the demand side, CMS assumes that as the leading edge of the baby-boom generation becomes eligible for Medicare, demand increases for health care, putting pressure on prices. CMS also expects that changes in private health insurance coverage will have a moderating effect on prices. In particular, health saving accounts, self-directed health plans, and

---

8. This series is labeled "HCFA Implicit Medical Price Deflator" in *National Health Care Expenditures Projections: 2005–2015*. Borger, et al., refer to the same series as a "PHC deflator."

9. Global Insight relied on CMS for its projection data. See *Projections of National Health Expenditures: Methodology and Model Specification*, p. 4.

10. Christine Borger, Sheila Smith, Christopher Truffer, Sean Keehan, Andrea Sisko, John Poisal, and M. Kent Clemens, "Health Spending Projections Through 2015: Changes on the Horizon," *Health Affairs*, Vol. 25. February 22, 2006, pp.w61–w73.

11. All these authors are with the CMS in varying capacities.

disease management programs should become more common. The impacts of these innovations, however, will be smaller than those seen from the managed-are revolution of the mid-1990s.[12]

On the supply side, the CMS assumes that health-care input prices will rise at rates above those seen in the past decade. Higher input prices should lead to higher health-care prices. The CMS projections also assume that "the diffusion of new medical innovations … continue to drive spending upward."[13]

(3) Public health insurance coverage: CMS projections incorporate Medicare and Medicaid policies in effect or slated to begin during the 2005–2015 projection period. These include the anticipated effects of Medicare Part D, expected to transfer significant spending from Medicaid and private sources to Medicare. As a result, public spending on health care is expected to faster than private spending. The Medicare Prescription Drug, Improvement and Modernization Act (MMA) of 2003 increased payments to managed-care plans. CMS projections assume that there is a resulting shift in enrollments from fee-for-service to managed-care plans.[14]

(4) Policies affecting physician payment growth rates: These projections also assume that Medicare's "Sustainable Growth Rate" system governing payment updates for physician services results in payment cuts in 2006 through 2013, when legislated cuts expire. Physician payments are expected to increase thereafter.

Assumptions for Expenditures per PRC in the Optimistic Scenario

For the optimistic scenario, we make the same assumptions on health-care cost growth by site-of-service, sectoral forces, public health insurance coverage, health-care input prices, and policies affecting physician payment growth rates.

We also look at additional trends that are likely to have a moderating effect on rising health-care cost growth. They include:

(1) Increases in Health Insurance Coverage: Increasing the number of insured can lead to lower prices because of volume discounting and the buying power enjoyed by third-party payers. More consumers will also face cost- containment strategies, such as denial of care.

(2) Growth in "Consumer-Directed" Health Insurance Plans: The optimistic scenario also includes the effects of the continued trend of increasing deductibles and other cost-sharing provisions that shift more of the health-care burden to employees. These dynamics create the incentives to use fewer health-care services, thereby dampening the growth in health expenditures.

(3) Innovations in Diagnosis, Treatment, and Prevention: As in the baseline scenario, we assume that innovations in health-care tend to increase spending and are likely to lead to higher site-of-service prices. But we do not expect them to have as great an impact on prices As a result, deploying new innovations will have less impact on prices.[15]

---

[12]. Borger, et al (2006), p.w65.
[13]. Borger, et al (2006), p.w65.
[14]. Borger, et al. (2006), p.w66.
49. Victor Fuchs and Alan M. Garber, "Medical Innovation: Promises & Pitfalls," *The Brookings Review* 21, no. 1 (2003).

Methodology

Disease Management Practices

Improved and more widespread adoption of disease management practices act to reduce the rate of future growth of health-care costs. Our optimistic scenario incorporates only moderate improvements in disease management practices. If greater advances in disease management practices are achieved, slower growth in health-care costs and treatment expenditures would be possible.

For example, more widespread breast self-examination or improved diagnostics would catch breast cancer at an earlier stage, when less-aggressive treatments are available, and reduce the growth in expenditures to treat patients. In the case of asthma (included in pulmonary conditions), improper management can lead to frequent hospitalizations and result in higher treatment expenditures. Improved disease management of diabetes can lessen the risk factors for developing cardiovascular disease and other conditions.

Health Information Technology

Surprisingly, firms within the health sector have been slow to adopt health information technologies (HIT), including electronic medical record systems (EMR). But providers, payers, and agencies will continue to install new health information capabilities and upgrade current capabilities. For example, according to a survey by the Medical Records Institute, many providers intend to add elements of EMR systems to their current HIT capabilities during the next four years. A significant numbers of providers are likely to implement EMR modules, such as data capture of lab results, progress notes, treatment warning, health screenings and post-visit patient education.[16] The CDC reports a 31 percent increase in the number of physicians' offices using full or partial EMR systems between 2001 and 2005.[17]

Disease-Specific Expenditure per PRC Projections

In order to project state expenditure per PRC, we created disease-specific expenditure growth rate index for onward twenty years. First, we calculated state average ratios of expenditure growth rate relative to the U.S., using CMS state personal health-care expenditure from 1993 to 2003. Then we generated disease-specific projections indexes of expenditure growth rate by multiplying these state average ratios by the U.S. annual growth rate of U.S. expenditure per PRC projections for twenty years onward. Finally, we obtained projections of expenditure per PRC by state, based on 2003 MEPS expenditure per PRC and the expenditure growth rate index.

Finally, we obtained disease-specific total expenditure projections by multiplying PRC by expenditure per PRC for onward twenty years.

---

[16]. MRinstitute. Eighth Annual Survey of Electronic Health Record Trends and Usage for 2006 www.medrecinst.com.

[17]. Centers for Disease Control and Prevention. See: www.cdc.gov/od/oc/media/pressrel/a060721.htm.

Part III: Historical Indirect Impacts (Forgone Economic Growth)

Good health is a vital component of individual well-being. But it also determines the economic contribution of an employee to a firm's success. When individuals suffer from chronic disease, the result is often diminished productivity, in addition to lost workdays. An ill employee who shows up for work (to avoid sick days, for example) may not perform well, a circumstance known as "presenteeism." Output loss due to presenteeism is immense; some literature suggests that for certain diseases, it can be up to fifteen times greater than for absenteeism, which is defined as work missed due to sick days etc.[18]

Caregivers also contribute to lost productivity through missed workdays and presenteeism. Currently, more than 20 million full-time employees provide care to others.[19] For this study, therefore, it is necessary to consider both employee groups for a more complete picture of the indirect impacts of chronic disease due to lost workdays and presenteeism.

We divided indirect impacts into four categories. First, any individual suffering or have suffered from any of the chronic diseases will have two main effects on work, lost workdays and presenteeism. Similarly, any person taking care of individuals with chronic disease will have an adverse impact on his/her work in form of the above-mentioned effects. Hence, in order to estimate indirect impacts we estimated all of the following:

A. Indirect impacts due to individual's (patients) lost workdays,
B. Indirect impacts due to individual's presenteeism,
C. Indirect impacts due to caregiver's lost workdays and
D. Indirect impacts due to caregiver's presenteeism.

In order to estimate indirect impacts, we use a wage-based and a nominal GDP-based(output) approach. For example, we multiply average wage with the number of lost workdays to estimate the wage-based indirect impact. Similarly, we use nominal GDP, for the GDP-based approach

We first calculate indirect impacts for diseases at the national level and then regional and state levels. In the following paragraphs, we outline the methodology to measure each of the sub-groups.

A. Methodology for individual's lost workdays:

First, we plan to measure the number of workdays missed by individuals who had contacted a chronic disease at some point. We compared the National Health Interview Survey(NHIS) and Medical Expenditure Panel Survey(MEPS) datasets to decide which one is more appropriate for this particular study. Although MEPS includes questions that can be related to our objective of study in this section, NHIS directly answers questions regarding number of lost workdays in a year. Hence, we decided to use the NHIS dataset.

[18] "The Hidden Competitive Edge - Employee Health and Productivity," (Newton, MA: Employers Health Coalition, 2000).
[19]. National Alliance for Caregiving and AARP, "Caregiving in the U.S." 2004.

Methodology

The NHIS public-use data is a nationally representative sample of the population in U.S. NHIS data has several components: the family core, household level, person level, sample-adult file, and the sample-child file. One adult from each family is randomly chosen to create the sample-adult file. The sample-adult file is representative of the adult population in the U.S. when appropriately weighted.

One of the questions in the sample adult file of the NHIS dataset asks each individual, "During the past twelve months, about how many days did you miss job or business due to illness or injury (not including maternity leave)?" This was one of the key questions for this section of the study. The NHIS sample adult file for 2003 survey was merged with family and person-level files to obtain more information about individuals and their families.

One of the limitations of the NHIS dataset is that it does not provide the exact number of lost workdays related to a particular disease. Hence, we had to use some proxy in this regard. We matched all employed individuals who ever had that particular chronic disease, whom we refer as Employed Population Reporting Condition (EPRC), with the number of lost workdays in past twelve months due to illness or injury.[20] We used this method to derive the number of lost workdays for each disease

EPRC for the U.S.*

Millions, 2003

Chronic Disease	EPRC
Cancer	5.9
Asthma	13.8
Diabetes	5.9
Hypertension	27.2
Heart Disease	9.5
Stroke	1.1
Emotional Disturbances	7.7

* Employed Population Reporting a Condition

Sources: NHIS, Milken Institute

Our next objective is to estimate indirect impacts of an individual's lost workdays for each of these chronic diseases. In order to do so, we multiply average wage per employee by the number of lost workdays by disease.

B. Methodology for individual's (EPRC) Presenteeism:

Once we estimated indirect impact of individual's lost workdays, we followed a 2004 study by Goetzel et al[21] to estimate an individual's (EPRC) presenteeism. They reported costs related to absenteeism and presenteeism (in addition to treatment costs) by disease. For example, the following table summarizes the findings from the Goetzel et al study.

[20] The only difference is in the case of "Emotional Disturbances," where we used the NHIS survey question "Have you seen/talk to a mental health professional in past 12 months?"

[21] RZ Goetzel et al., "Health, Absence, Disability, and Presenteeism Cost Estimates of Certain Physical and Mental Health Conditions Affecting U.S. Employers," *Journal of Occupational and Environmental Medicine* 46 (2004).

Costs Related to Absenteeism and Presenteeism

Per Employee, Annual

Chronic Disease	Absenteeism	Presenteeism
Cancer	4.5	75.7
Asthma	2.1	72.2
Respiratory Infections	27.5	33.3
Diabetes	19.2	158.8
Hypertension	46.7	246.7
Heart Disease	19.2	70.5
Emotional Disturbances	33.4	246.0

Source: Journal of Occupational and Environmental Medicine, 2004

We used disease-specific ratios of presenteeism to absenteeism from their study and multiplied by our estimates from individual's lost workdays to derive indirect impacts due to individual's presenteeism.

C. Methodology for caregiver's lost workdays:

The National Alliance for Caregiving and AARP[22] measured that total number of caregivers in the U.S. is 44.4 million (39 percent of men and 61 percent of women). Out of this, 60 percent of men and 41 percent of women are full-time employed.

Caregivers in the U.S.

Millions

	Gender	
Caregivers	Male	Female
Total	17.3	27.1
Full-Time Employed	10.4	11.1

Source: NAC and AARP, 2004

A study by Metlife (2006) shows that 10 percent of men and 18 percent of women on average miss 9.0 and 24.75 workdays, respectively, for caregiving purposes. In order to get the number of lost workdays for caregivers at the national level, we used the above information.

Next we allocated caregivers' lost workdays by using disease-specific percentages of lost workdays to total lost workdays due to all type of illness or injury (from individual's lost workdays).

D. Methodology for caregiver's presenteeism.

In order to estimate caregivers' presenteeism, we first calculated employed caregivers by condition (ECC). For example, we found that EPRC for cancer in 2003 (from individual lost workdays) was 5.92 million, which accounted for 3.5 percent of all employed population in that

[22] Russonello & Stewart Belden, "Caregiving in the U.S.," (National Alliance for Caregiving and AARP, 2004).

year (from NHIS). Following that, we allocated 3.5 percent of all full-time employed caregivers (21.5 million) to cancer (0.77 million).

ECC for the U.S.*

Millions, 2003

Chronic Disease	ECC
Cancer	0.77
Asthma	1.78
Diabetes	0.76
Hypertension	3.52
Heart Disease	1.23
Stroke	0.14
Emotional Disturbances	1.00

* Employed Caregivers by Condition

Sources: NAC, Milken Institute

Next, we calculated ECC adjusted individual presenteeism. For cancer, wage-based individual presenteeism was $103.71 billion, after adjusting for ECC, it turned out to be $13.42 billion. Next, following a report by Levy,[23] we then allocated 75 percent of ECC adjusted individual's presenteeism as caregivers' presenteeism. For cancer, 75 percent of $13.42 billion is $10.06 billion. We followed the same methodology to estimate caregiver presenteeism for other diseases.

Methodology for estimating indirect impacts for different cancer type

In estimating indirect impacts for diseases as outlined in earlier paragraphs, we started with the NHIS dataset. NHIS dataset however, does not provide similar questions whether an individual ever had a particular type of cancer (breast, colon, lung, prostate etc.). Hence, indirect impacts for different types of cancer were estimated by using expenditure shares of different types of cancer from Historical Direct Cost estimation. For example, breast cancer accounted for 11 percent of total expenditure on cancer; colon (8 percent), lung cancer (13 percent), and prostate cancer (9 percent). Other types of cancer constituted the rest, 59 percent.

Methodologies for regional and state level indirect impacts

Our next step is to determine lost workdays related to major chronic diseases, broken down by census regions. We controlled for inter-regional variations by taking regional shares for employed population reporting condition (EPRC) and lost workdays per EPRC for each disease and averaged them over three years (2003-2005). Then we scaled them up to 2003 national values to obtain revised EPRC and revised lost workdays per EPRC. The revised lost workdays are obtained by multiplying revised EPRC with revised lost workdays per EPRC.

One of the problems with the NHIS public-use data is that it does not provide information about state identifiers. It is difficult to obtain detailed information about each state from the NHIS confidential datasets, managed by National Center for Health Statistics, a division of CDC. In

[23] David Levy, "Presenteeism: A Method for Assessing the Extent of Family Caregivers in the Workplace," (American Association for Caregiver Education, 2003). and David Levy, "Presenteeism: A Method for Assessing the Extent of Family Caregivers in the Workplace and Their Financial Impact," (American Association for Caregiver Education, 2007).

order to estimate indirect impacts by state, we used state-level PRC shares (obtained from the Historical Direct Cost Estimation) for each disease. Example: if PRC share (out of national) for a disease is 5 percent for a particular state, we allocated 5 percent of the national indirect impact as that state's indirect impact. State level impacts for different types of cancer were obtained from using expenditure shares of each cancer type to that state's total cancer expenditure.

Part IV: Projecting Avoidable Indirect Impacts (Forgone Economic Growth)

A. Baseline and Optimistic Projections

In this part of the study, we extend our findings from the previous section to project future indirect impacts. We will project indirect impacts under two alternative scenarios-the baseline and the optimistic. The avoidable indirect economic impact is defined as the difference between the baseline and optimistic projections

Baseline Scenario

In developing baseline and optimistic scenarios of future indirect impacts, we first projected future path of employed population reporting condition (EPRC) and employed caregivers by condition (ECC) using projections of employment (from Economy.com and U.S. Census) and population reporting condition (PRC) (from Projecting Avoidable Direct Costs).

Next, we use employment and population projections to calculate employment-to-population ratios (population is defined as 16 years and older). Next ratio for every year is divided by that for 2003, to we build an **E/P index**. For example, the E/P index for 2004 was derived by dividing the 2004 employment-to-population (0.58) by the 2003 ratio (0.58).

Projections of Employment and Population

Year	Employment (Millions)	Population* (Millions)	Employment/ Population	E/P Index
2003	130.0	225.2	0.58	1.000
2004	131.4	227.7	0.58	1.000
2005	133.5	230.3	0.58	1.004
2006	135.4	233.0	0.58	1.006
2007	136.8	235.7	0.58	1.005
2008	138.3	238.2	0.58	1.006
2009	140.1	240.6	0.58	1.008
2010	142.0	242.9	0.58	1.013
2011	144.0	245.1	0.59	1.017
2012	145.9	247.3	0.59	1.022
2013	147.8	249.3	0.59	1.027
2014	149.8	251.3	0.60	1.032
2015	151.7	253.4	0.60	1.037
2016	153.7	255.4	0.60	1.042
2017	155.7	257.6	0.60	1.047
2018	157.6	259.8	0.61	1.051
2019	159.6	261.9	0.61	1.056
2020	161.6	264.1	0.61	1.060
2021	163.5	266.3	0.61	1.064
2022	165.4	268.5	0.62	1.067
2023	167.3	270.7	0.62	1.070

* Adult Population is defined as 16 years and over

Sources: BLS, U.S Census, Economy.com, Milken Institute

We next create a **baseline PRC index** for each disease. This is built by dividing baseline PRC (obtained from "Projecting Avoidable Direct Costs") for every year by baseline PRC for 2003. The following table provides an example: PRC index for cancer. The index reading for 2004 (1.03) is derived by dividing 2004 PRC (10.93 million) by 2003 PRC (10.58 million).

Cancer

Projection of Lost Workdays

Year	PRC (Millions)	PRC Index	E/P-PRC Index*	EPRC (Millions)	Lost Workdays (Millions)
2003	10.58	1.00	1.00	5.92	60.14
2004	10.93	1.03	1.03	6.11	62.09
2005	11.25	1.06	1.07	6.36	64.59
2006	11.61	1.10	1.10	6.58	66.81
2007	12.00	1.13	1.14	6.79	68.99
2008	12.35	1.17	1.17	6.99	71.04
2009	12.70	1.20	1.21	7.21	73.22
2010	13.03	1.23	1.25	7.43	75.44
2011	13.36	1.26	1.28	7.65	77.73
2012	13.72	1.30	1.33	7.90	80.22
2013	14.06	1.33	1.36	8.13	82.58
2014	14.39	1.36	1.40	8.36	84.91
2015	14.71	1.39	1.44	8.59	87.22
2016	15.01	1.42	1.48	8.81	89.48
2017	15.33	1.45	1.52	9.03	91.73
2018	15.64	1.48	1.55	9.26	94.05
2019	15.97	1.51	1.59	9.49	96.39
2020	16.30	1.54	1.63	9.73	98.81
2021	16.62	1.57	1.67	9.95	101.11
2022	16.95	1.60	1.71	10.18	103.44
2023	17.28	1.63	1.75	10.41	105.74

* E/P-PRC Index was created by multiplying the E/P Index with the PRC Index

Sources: BLS, U.S Census, Economy.com, Milken Institute

We multiply the E/P index by the PRC index to create an **E/P-PRC index**, also shown in the above table. This index is scaled to the 2003 EPRC to obtain projections of EPRC by disease. For example, in 2003, cancer EPRC totaled 5.92 million. Hence, each year's EPRC is multiplied by 5.92 million to obtain cancer projections of EPRC through 2023.[24]

Baseline EPRC are converted into lost workdays and presenteeism for both individuals and caregivers consistent with the methodology used to estimate the indirect impacts (Historical Indirect Impact).

We then use projections of wages and nominal GDP, respectively, to obtain wage- and GDP-based projections of indirect impact for the baseline scenario.

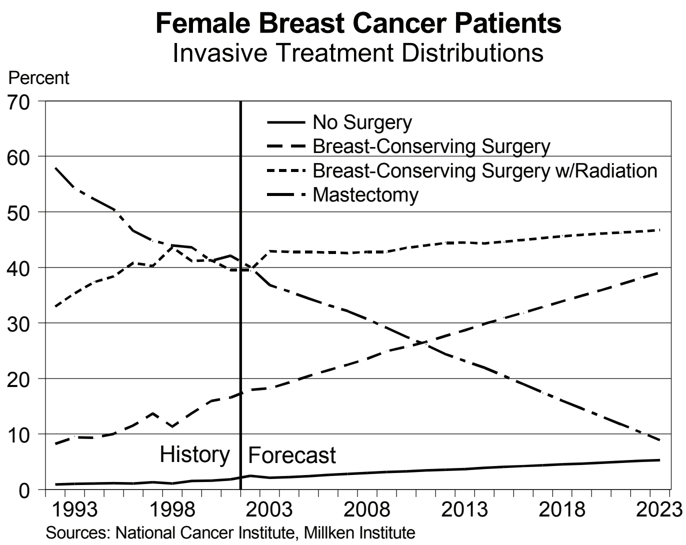
Optimistic Scenario

In this scenario, the indirect economic impacts of lost workdays are calculated as they were for the baseline scenario, using optimistic PRC figures from the "Projecting Avoidable Direct Costs" section.

[24] We followed the same methodology to calculate projections of ECC by disease.

However, we don't just want to incorporate optimistic PRC. We also want to include changes in treatment that will reduce presenteeism through less-invasive treatments and lower side effects. This reduction will ultimately affect the indirect impact ratio of presenteeism to lost workdays.

We assumed that the maximum effect on presenteeism will be for cancer. We followed a report by the National Cancer Institute[25] on percentages of no surgery, BCS (breast-conserving surgery) with radiation, BCS without radiation, and mastectomy from 1992-2002 for breast cancer patients. This report on breast cancer is one of the best available information and can be used as a proxy to measure the relative invasiveness of treatment options for other diseases. Keeping this in mind, we started by ranking the above four categories according to least invasive to most: (1) no surgery; (2) breast-conserving surgery (BCS) without radiation; (3) BCS with radiation; and (4) mastectomy. Ranking the four options, we project each out through 2023.



We next want to project future ratio of presenteeism to lost workdays. We assumed it is affected by all four treatment options for cancer. However, we also assume that certain treatments will have a greater effect on presenteeism: (1) no surgery (highest); (2) BCS with radiation; (3) mastectomy; and (4) BCS without radiation (lowest). Since we are not definite about the magnitude of variation in presenteeism among the first three categories (no surgery, BCS without radiation and mastectomy), we used equal weights for these three (0.3 each) and 0.1 for BCS without radiation. Next we created a weighted index[26] and deflated the 2003(baseline) presenteeism to lost workdays impact ratio by that. The following table shows presenteeism to lost workdays impact ratio for cancer.

[25] "Cancer Trends Progress Report: 2005." See: www.cancer.gov.

[26] For each of these series, we used 2003 as base year.

Cancer

Presenteeism to Lost Workdays

Year	Presenteeism / Lost Workdays	Absolute Change
2003	16.95	–
2004	16.64	-0.308
2005	16.30	-0.343
2006	15.95	-0.343
2007	15.62	-0.340
2008	15.33	-0.284
2009	15.04	-0.294
2010	14.86	-0.177
2011	14.63	-0.229
2012	14.46	-0.166
2013	14.30	-0.169
2014	13.96	-0.333
2015	13.76	-0.204
2016	13.57	-0.190
2017	13.39	-0.182
2018	13.21	-0.178
2019	13.03	-0.179
2020	12.85	-0.177
2021	12.67	-0.182
2022	12.50	-0.173
2023	12.33	-0.171

Source: Milken Institute

For other chronic diseases, we project the indirect impact ratio through an ordinal ranking, by disease, and try to ascertain the relative effects of the four treatment options on each. The rationale behind such a ranking is borrowed in part from the number of ongoing clinical trials.

The next table gives totals for ongoing clinical trials, as of early 2007. Cancer is the subject of most trials. We assume that more clinical trials will lead to less invasive treatment options.

Clinical Trials by Disease

Chronic Disease	Total
Breast Cancer	543
Colon Cancer	337
Lung Cancer	441
Prostate Cancer	257
Heart Disease*	1,532
Diabetes	447
Pulmonary Conditions	145
Depression	297

* Including Hypertension and Stroke

Source: ClinicalTrials.gov

We also assume that less invasive treatment options will affect future presenteeism, another factor in building the ordinal ranking. The concept is summarized in the following table:

Effect of Invasive Treatments on Presenteeism by Disease

Chronic Disease	No Surgery	BCS* (without radiation)	Mastectomy	BCS* (with radiation)	Percent Compared to Cancer
Cancer	X	X	X	X	100
Heart Disease	X	X	X		60
Diabetes	X	X			35
Stroke	X				25
Asthma	X				20
Emotional Disturbances	X				15
Hypertension	X				10

* Breast-Conserving Surgery

Source: Milken Institute

Heart disease is affected by drugs (to relate to the above table, no surgery), part-surgery (BCS without radiation), and full-surgery (mastectomy). So using similar weights, as in cancer, we would assume that the change in presenteeism for heart disease to lost workdays impact for every year is proportional to yearly changes in that for cancer. Thus, if in 2003-2004, the absolute change in presenteeism to lost workdays impact for cancer was (-0.31), then we would assume that for the same period, similar change for heart disease would be 60 percent[27] of that. Next we scaled it down by 2003's presenteeism to lost work days impact ratio of heart disease to cancer.[28] We followed this methodology to get the presenteeism to lost workdays impact ratio for heart disease as in the following table:

[27] Following the above logic, heart disease should contribute to 70 percent of the change as in cancer. However, we used 60 percent to allow for any additional effect specific to heart disease.

[28] Hence, the final change in presenteeism to lost work days impact ratio for heart disease from 2003 to 2004 will be (-0.31X 0.60) X (3.63/16.95).

Heart Disease
Presenteeism to Lost Workdays

Year	Presenteeism / Lost Workdays	Absolute Change
2003	3.63	-
2004	3.59	-0.046
2005	3.54	-0.051
2006	3.49	-0.052
2007	3.43	-0.051
2008	3.39	-0.043
2009	3.35	-0.044
2010	3.32	-0.027
2011	3.29	-0.034
2012	3.26	-0.025
2013	3.24	-0.025
2014	3.19	-0.050
2015	3.16	-0.031
2016	3.13	-0.029
2017	3.10	-0.027
2018	3.07	-0.027
2019	3.05	-0.027
2020	3.02	-0.027
2021	2.99	-0.027
2022	2.97	-0.026
2023	2.94	-0.026

Source: Milken Institute

Following similar logic, we applied 35 percent of changes in cancer for diabetes. For the other diseases, it is only affected by drugs (no surgery). But in order to bring in some variation, we assumed, stroke will have 25 percent of the change, followed by asthma (20 percent), emotional disturbances (15 percent) and hypertension (10 percent).

B. Projections of Avoidable Indirect Impacts

The avoidable indirect economic impact is defined as the difference between the baseline and optimistic projections.

Part V. Forgone Economic Growth and Intergenerational Impacts

Background

While the contemporaneous impacts of lost workdays, wages, and productivity due to chronic disease are substantial, the longer term or intergenerational impacts on economic growth are likely to be of a far greater magnitude. Yet there has been little research or attempt to quantify the impact of poor health (chronic disease) on human and physical capital formation and the restrictions imposed on long-term economic growth. We develop a methodology and build a model allowing us to demonstrate and forecast the impacts of chronic diseases for all fifty states. We do not believe that this has ever been attempted.

Since the early 1990s, there has been a renewed emphasis on studying ways to identify the determinants of economic growth. Most of these efforts fall under the endogenous growth theory umbrella. Endogenous growth theory is based upon the observation that the factors that influence economic performance are determined within the system and simultaneously interact with each other. Many variables and model specifications have been attempted, but only a few have been found to be statistically significant in explaining growth.[29]

Human capital's critical role is now widely recognized among economists. Dynamic economic growth depends on the stock of human capital and continued investment in education, new work-based learning and training procedures (flow), and greater levels of health. In advanced economies, if new investments in human capital fall below the advanced-economy average, economic growth and per capita income advances will lag behind and risk tipping into a downward spiral. Better health leads to greater investment in education, resulting in ever higher levels of human capital. This causes wealth to increase and leads to a virtuous cycle of economic growth. At the macroeconomic level, increased health, lower chronic disease and improved life expectancy raise the rate of return to a variety of investments. The result is faster capital accumulation and a tipping point created that ignites an explosion in knowledge and technology that can be harnessed to improve economic growth further.

Good health increases the rate of return to investments in education. Studies show that children who are well nourished, energetic, and spirited will gain more from incremental education than children who are malnourished and tormented by the incapacitating effects of chronic disease. Another benefit of good health is that it tends to make people more creative. Similar to a person being more efficient in producing goods or services in the workplace, healthy individuals are more likely efficient in creating new knowledge. This leads to expanding the research-production-possibilities frontier and improves a nation's competitiveness in the long term. Better health also improves a person's ability to cope with stress and adapt to rapid, sometimes stressful technological change.

When an income earner attempts to determine the money that the household will consume, save, or invest in human or physical capital, he or she is making decisions on how to maximize intergenerational wealth transfers. The higher the income earner's human capital, the greater the probability that he or she will invest heavily in their children's and grandchildren's education. The high correlation between adult income and health is largely the result of past

---

[29] Guillem López-Casasnovas, Berta Rivera, and Currais Luis, *Health and Economic Growth: Findings and Policy Implications.* (Cambridge: The MIT Press, 2005).

intergenerational dynamics, namely, the impact of child health, itself determined by family endowments, on future adult health, education, and income.

One must be careful not to see investments in health in the context of higher health-care expenditures. Investing in health and health-care spending are mutually exclusive activities. Investing in health requires a broad-based strategy attempting to identify health population targets in order to close the gap between health prevention, early diagnosis, and treatment in the real world, with leading efficacy in the laboratory of medical research. An under-investment in health leads to an under-investment in human and physical capital, as well as lower economic growth and wealth.

Production Function Explanation

We deploy an empirical model-building strategy that permits us to estimate the effects of health on long-term economic growth by incorporating an aggregate production function methodology that treats health as both a separate factor of production and a dynamic interaction with the accumulation of human and physical capital. A production function explains measures of output (gross domestic product by state) as a function of inputs (physical and human capital, health and technological change). The methodology permits an endogenous feedback mechanism but simplifies the complex estimation procedures of a fully endogenous framework. This approach allows prediction of lost or potential gained economic growth due to chronic disease on the current generation (twenty years ahead) and the intergenerational impacts (forty years ahead).

A production function is an equation that describes how factor inputs translate to output (real GDP of states in our framework.) It relates the technology involved in the process. The regression coefficients for each specific factor input relay the strength and magnitude of the relationship between that specific factor and output. A positive coefficient on factor X for example, indicates that increasing X will increase output.

This aggregate production function approach involves overcoming some challenges stemming from the intangible nature of some of the components of human capital and health. Similar to human capital in that educational attainment and investments in private and public workforce training are used as proxies in measuring it, health status is essentially a non-observable variable requiring imperfect proxies to be developed.

Our approach limits these challenges in the empirical analysis. We estimate a variant of a Cobb-Douglas production function specification. However, a procedure will be developed that permits cross-regional variation in technological progress so that a constant rate of technology diffusion isn't imposed. This will result in state-specific intercept terms.

In the log linear model formulation, the estimated coefficients will be elasticities. This allows us to state the elasticity of income (output) with respect to health. Additionally, due to the interaction between health and physical and human capital, we will be able to capture the endogenous impact of health on the entire system that accounts for economic growth, further magnifying its effect. The results of the production function can be interpreted as relatively short-term elasticities when compared to the cross-sectional regressions used in the intergenerational analysis, which resemble more long-term elasticities. The reasoning behind this differentiation is that there is more variation between states than over the time period used in the production function.

Methodology

A survey of comparable literature reveals the concentration of research on links between economic growth and health and education. Controlling for basic traits like geographic location, population growth, secondary schooling, and institutional characteristics like openness and government savings, research in this field have pointed to the relationship between life expectancy change and economic productivity. For example, a one-year change in life expectancy at birth leads to a 4 percent boost in productivity. This result from Bloom, Canning and Sevilla's seminal "World Development 2003" paper is comparable with established results. However, one must consider the limitations of applying their results to a market like the United States.

One caveat is the time period under examination. Most papers have looked at the 1960s through early 1990s. This period was marked by recessions and high inflation in many countries. The latter portion was also the start of a thrust toward capital stock accumulation, particularly software, in advanced countries. The 1980s and1990s also benefited from the introduction of innovations in the antibiotics market. However, most papers have few observations in this time period (i.e., decennial cross section), and this may not enable them to fully capture the relationship between health and growth. Some countries would have had drug entry much earlier, depending on pharmaceutical industry penetration and market appeal. With few observations and a lack of control over intellectual property rights and regimes, some countries may have been ruled as outliers when in fact when we examine alike countries, or states in our case, with the same governmental regime on intellectual property protection, one can more safely attribute increases in life expectancy and increases in investments to health to GDP and productivity.

The issue of countries represented in the historical panel dataset is of particular concern. Because existing literature has focused on non-OECD countries, the variables used are generally not applicable to our research. Developing countries have a particular set of concerns. Their target on increasing adult male survivorship, increasing immunizations for measles, and public access to sanitary water supply, for example, is not especially relevant to the United States in the latter half of the 20th century. Therefore though we can pattern our production function in the same manner, using a log-linear model, we must carefully consider the type of right hand side variables to include.

Data Examined

(1) Life Expectancy
The literature has used both life expectancy and male survival rates to proxy for improvements in health. These studies have mainly focused on developing countries where improvement to health status first impacts males and has a significant impact on survivorship of all individuals. Considering that our research examines the United States from the late 1960s to the present day, the use of male survivorship is outdated. Rather, we use life expectancy both at birth and at age 65, as well as mortality. Greater life expectancy and better health status are usually synonymous. Recent health literature points to use of life expectancy at 65 years old to decipher the trend in chronic health conditions. Life expectancy at 65 a good measure of the cumulative investments to diet, nutrition, and lifestyle factors, as well an innovative way to evaluate trends in chronic health, and unlike life expectancy at birth, it is not as confounded. Life expectancy at birth is complicated by the high rates of infant mortality and other extenuating circumstances that might lead older women to try for difficult births, particularly in more affluent, technologically advanced countries. Life expectancy at 65 seems to be the most direct proxy for chronic disease.

These life expectancy numbers were taken from the NCHS and derived from mortality data extending back to the late 1930s. Decennial state information was computed by the CDC from mortality information from the late 1930s to the 1990s in conjunction with the U.S. Census population information. However, the latest 2003 figures were taken from publicly unreleased CDC mortality data along with the U.S. Census population data for 2003.[30] Given the decennial regional data and the associated annual national figures, we are able to compute, using a geometric series, each state's life expectancy number.

(2) Human Capital
An important but indirect contribution of a healthier aging and employed population is the increased quality of the work force, measured by the amount of formal schooling attained. This is especially relevant, considering that the baby boomer generation has generally been acknowledged as the most well educated work force. While the U.S. workforce is slowly undergoing potentially pessimistic changes in the amount of increased diversity in the labor force and the retirement of the boomers, alleviating chronic health in the next few years could ameliorate this labor supply growing pain.

Average educational attainment of the population was explored for use as a proxy for the human capital variable. While this is an imperfect measure due to non-degreed improvements in human capital acquired through the workplace or in non-accredited training programs, there is generally a very strong correlation between initial educational attainment and subsequent investments. Because we are investigating the returns to human capital at the state level in an advanced economy, it necessary to explore estimating the separate impacts by degree attainment.

Proxy variables are typically chosen that mimic the unobservable variable as closely as possible, but these can't exhibit the full range of conditions or are linked to a single facet of health. In essence, this creates measurement error in the explanatory variables due to variations in heterogeneity and causal feedback between health, human and physical capital, and productivity or technological change.

(3) Capital Stock
Decreases in morbidity, and increases in longevity, create the need for individuals to save for their retirement. Increased savings leads to greater private investment in plant, equipment and technology, and are discernable in public capital infrastructure investment improvements. As physical capital accumulates, it increases aggregate efficiency, directly impacting a region's economic output per capita.

(4) Dependent Population
Within the model, we accounted for the changing number of the dependent population on each state's productivity. States with disproportionate shares of the young are likely to adversely affect output levels in a production function formulation. These affects will be severe in states such as Utah.

In advanced economies, reducing chronic disease and morbidity may cause the retirement age to increase. Extending the productive capacity of individual human capital by lengthening their active workforce participation could have a huge collective impact on improving economic growth.

[30] The 2003 life-expectancy-at-65 numbers were computed by Dr. David Solet at the Planning and Evaluation Department of the Seattle & King County Public Health Unit.

Methodology

Production Function Used

We use the following production function to describe how state output is attained.

The reduced form equation is presented first in equation (1).

$$Y = F(K, L, H, YD, E) \qquad (1)$$

The expanded production function is shown below.

$$Y = A\ K^{\alpha}\ L^{\beta}\ H^{\delta}\ YD^{\lambda}\ E^{\gamma} \qquad (2)$$

where Y is the output or gross domestic product by state (GDP); A represents total factor productivity; K is physical capital and is composed into a combination of private and public non-residential structures, and software and equipment; L is the conventional labor force; and human capital is characterized by the percent of the state population with a bachelor's degree and above (E); while health, H, is where we utilize life expectancy at age 65. For health, we considered life expectancy at birth, as well as overall mortality rates for each state but we felt that the variable we ended up using, life expectancy at 65, captured the essence of our research more directly. Variable YD represents the fraction of the population between the ages of 0 and 16.

We captured the flavor of the Bloom, Canning, and Sevilla (2004) model but address their concerns about model flexibility and lack of data issues by incorporating into our model output and by obtaining life expectancies at birth and at age 65, as well as overall state mortality rates to proxy for health (H = life expectancy at birth, life expectancy at age 65, and mortality).

Taking logs of the aggregate production function, we derive an equation for the log of the output in state *i* at time *t*.

$$y_{it} = a_{it} + \alpha\, k_{it} + \beta\, l_{it} + \delta\, h_{it} + \lambda\, yd_{it} + \gamma\, e_{it} + \varepsilon_{it} \qquad (3)$$

where the lowercase represent the logs respectively. We will then estimate the values for the following coefficients: $\alpha, \beta, \delta, \lambda, \gamma$.

The production function approach utilizes a straightforward equation of the natural logs of inputs in order to produce results in terms of elasticity to output. A fixed-effects model was developed to incorporate state variation. Time dummies were considered for this estimation. A balanced panel dataset was assembled from 1970 to 2003, which included unadjusted labor force, physical capital stock, life expectancy at 65, dependent population per capita, and a measurement that captures human capital formation—the percentage of the population with a bachelor's degree and above. Results are presented in the table below.

Production Function Results
Dependent variable: Log (Real GDP by state)

Variables	Coefficient
Log (Life expectancy at 65)	0.258* (2.05)
Log (Bachelor's degree)	0.506** (19.31)
Log (Unadjusted labor force)	0.7499** (26.17)
Log (Capital stock)	0.196** (14.84)
Log (Young dependent per capita)	-0.311** (-7.09)

*significant at the 5% level
**significant at the 1% level
Source: Milken Institute

The results relay the beneficial aspects of health and education toward output. They confirm previously established literature results but also point more specifically to the use of more sophisticated variables to describe the next step for most developed country estimation. Life expectancy at 65, or the cumulative benefits from investing in lifelong health, contributes significantly to increasing productivity. As previously discussed, the numbers of young dependents will also influence output. Their exclusion from the labor force gives us the predicted negative coefficient in the production function estimation. Labor force and capital stock are inputs in any GDP equation and as predicted, they are positive and highly significant. However, the production function estimation is only the first step of the intergenerational analysis. Regression results above demonstrate the factor inputs' short-term elasticities to output. The next step is to forecast how increasing current investments in education and health will demonstrably impact future investment decisions and feed back to state productivity.

Variables Used (Intergenerational Analysis)

An innovation with the long-term economic impact model is the incorporation of intergenerational contributions to a state's output, given the health investments of individuals over a forty-year time span. This translates to a more sophisticated econometric model which utilizes dynamic reaction functions to explain how investments in health have a positive spillover into investments in human capital, capital stock and augments the labor force.

For this section of the analysis, variables used in the production function were forecasted to 2050 for all states (excluding the District of Columbia) to compute both baseline and optimistic scenarios. The U.S. computation is the sum of states. The two forecasted scenarios allow us to compare the percentage difference in state output.

Methodology

A. Baseline Scenario

(1) Gross Domestic Product

We utilized Global Insight's smoothed-out real GDP by state series and extended the horizon back to 1969. The BEA's new NAICS code-based GDP extend back to 1989. We utilized the BEA's discontinued SIC codes-based nominal GDP series and deflated it by using the 2000 chain-weighted GDP deflator. These two series were spliced together in 1988. This gave us our continuous, long-term dependent variable.

(2) Life Expectancy at 65

In 1966, the U.S. life expectancy at 65 was 14.60 life years remaining. It increased to 18.40 by 2003. The difference was roughly 3.8 life years: averaging 0.1 years at age 65 for every passage of a calendar year. The baseline took this historical growth into account and we figured that by 2023, the baseline forecast should reveal a gain of roughly 2.18 years. By 2050, there should be an additional gain of 2.945 life years from the 2023 marker. This would bring the 2050 baseline figure to 23.53 life years remaining at age 65. We assumed a simple constant annual growth to populate the series in between the 2023 and 2050 markers.

(3) Bachelor's Degrees

Human capital development is captured by the percentage of the population who have a high school degree and above, and the percentage of the population with a bachelor's degree and above. The information available from the Census Bureau's Decennial Census of Population is smoothed out using the annual national pattern for each respective type of degree, high school or bachelor's, taken from the Current Population Survey's historical tables. The state data is available from 1988 to 2003; however, this has a rather high standard error. A more accurate accounting would not have been so erratic, but utilizing the data available, we employ a five-year moving average to smooth out this series. For the data from 1970 to 1987, we have decennial state information and a complete annual U.S. series from the census bureau. We use the U.S. series as a pattern for the state series construction. So data from 1970 to 1987 was smoothed out using the U.S. bachelor degree education data, while the real data from 1988 to 2003 was smoothed out with a five-year moving average to minimize the impact of erratic yearly deviations.

For bachelor's degree as a percentage of the population, the construction of forecasted variables involves running the regression of state degrees divided by the national numbers against state dummies and a time trend. Incorporating the Cheeseman Day and Bauman (U.S. Census Population Paper No. 43), we calculate that national figures will increase 0.208 percent each year from 2003 to 2028 according to the paper, which accounts for future changes in race, as well as sex, nativity, and age. We keep this rate constant and extend the period of growth to 2050. Given the creation of this series of national percentages, we can now multiply this against the state share ratio and add year and year-squared variables to get unique state percentages of bachelor's degree and above for each year from 2004 to 2050.

(4) Capital Stock

Two different approaches were applied to derive the amount of state-level private and public capital assets. First, private fixed assets: both structures, and equipment and software, were compiled through assignation of state shares to national levels of stock. The state shares are deduced through construction of state wage shares. Wage data from the BEA gives a reliable indicator of level of earnings available for investment in private assets by state. National nominal figures for private, non-residential structures, and software and equipment, respectively, are adjusted with their associated investment price index. A private, non-residential structure price index was applied to the national stock of private, non-residential structures and likewise for equipment and software.

Second, public fixed assets were estimated using state and local government expenditures in the form of capital outlays and construction, which are available at the state and local government level from the U.S. Census beginning in 1988. This was eventually used to construct state shares of public fixed assets. State growth of capital outlays was used to augment the state and local government dataset for both capital and construction back to 1957. When recent construction state-level growth rates were available, they were utilized to reconstruct the missing state and local government construction data. Missing data in in-between years and outliers were imputed assuming a simple linear relationship between adjoining years. Like national nominal figures for private, non-residential structures, the national nominal figures for public structures were adjusted using a government investment price index. A similar technique was applied with public equipment and software.

In the final regressions, we used the overall capital stock to investigate the impact of physical capital on state productivity. However, given the technological spur in the past decade from Silicon Valley and the tech industry, it is appropriate to attribute that not only will the impact of software and equipment be positive, but that it will be greater in magnitude than structure's contribution to state productivity. This should be especially true in California, for example. A caveat is the inability to disseminate through this approach the difference in contributions, of government's investment in software and equipment stock versus the private sector's investment.

The capital stock variable is developed by first investigating three of its properties: software, equipment and structures. The U.S. data is available from Global Insight and is composed of several smaller components like commercial equipment, industrial equipment, information equipment, miscellaneous equipment, non-residential computer equipment, non-residential miscellaneous other equipment, and non-residential other equipment. The structures variable from the Global Insight forecast bank represents simply the real net stock of non residential building in billions of chained 2000 dollars. We then sum these subcomponents to form a U.S. capital stock number and divide by Economy.com's projected employment figures. We calculated yearly growth of U.S. capital stock per employee. We then projected to 2050, using the last year of data growth rates. The growth rates were then applied to the state capital stock total per employee from 2003 onward to gather an estimate of each state's capital stock variable from 2004 to 2050 after adjusting for unique projected state employment.

(5) Young Dependents per Capita

Annual estimates of the population by age at the state level are available from the U.S. Census from 1970 to 2004. Data is categorized by five-year age cohorts. By data restriction, while adhering as close as possible to the definition of working age population, the dependent populations will comprise all ages 14 and below, and over 64, while the working age will comprise all ages between 15 and 64 years. Total population figures are available from Bureau of Census Current Population Reports from 1940 to 2005.

Economy.com has the population projections by one-year cohort in its demographic projections bank which is based on the Census Bureau. We sum up ages 0–16 by state and divide by the total state population for each year. In the last year, we take a look at each state's growth rate and keep the same growth till 2050. The baseline forecast for young dependents per capita and the optimistic forecast are the same figures. When we compute the impact of health to real GDP by state, we try to hold everything else that is not involved in the reaction functions constant. One might argue that greater education will lower birth rates but that argument is more acceptable in developing countries.

(6) Unadjusted Labor Force

The average of the monthly labor force numbers were taken from the BLS from 1976 to 2006. Labor force represents the population over 16 actively looking for a job. This includes those already employed, as well as those unemployed and searching for work. The employed population constitutes a far larger percentage of the total labor force statistic. To derive 1969–1975 data, a labor force proxy of employment per capita was created. We applied the growth rate from this series to labor force to assemble the complete series.

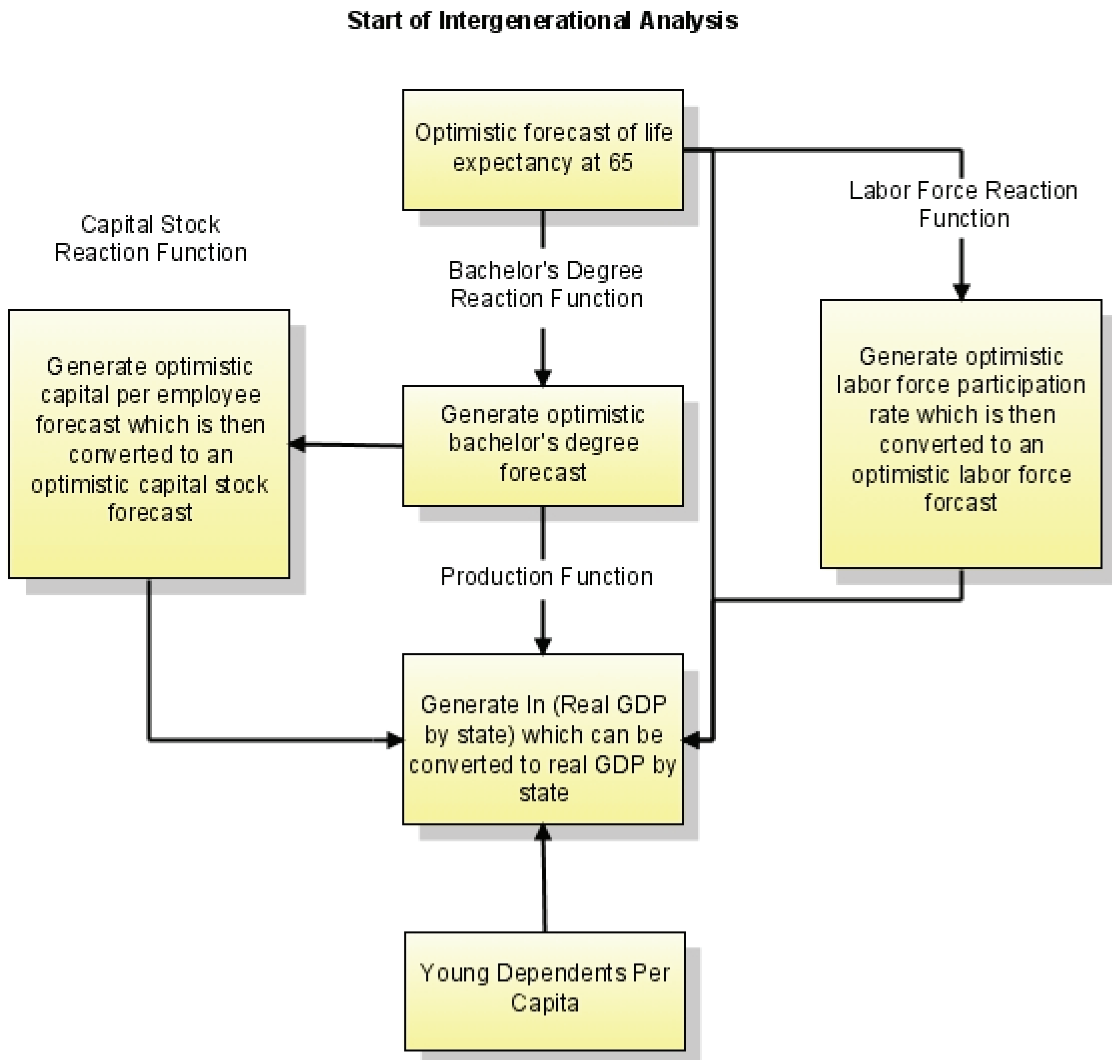
For labor force and capital stock baseline forecasts, we first create a U.S. series and utilize the growth rate in the last period of observation in year 2030 to propagate the individual state series to 2050.

We cannot simply project labor force without accounting for changes in the population. We therefore take historical BLS labor force figures to compute the current labor-force participation rate. The next step is to forecast the labor-force participation rate and multiply the yearly ratio by the census bureau's population projections to derive annual labor-force numbers. Labor-force projection numbers at the aggregate level were available through Global Insight's civilian labor force, with adjustment for 2000 Census from the BLS. Information regarding population over 16 was available through Economy.com in its demographic population projections by one-year cohort. Although U.S. labor-force participation rates were eventually growing at .33 percent, ultimately each state's labor-force numbers will be different depending on what their initial labor force participation rate was in 2003, as well as on that state's population projection over 16.

B. Optimistic Scenario

The flow chart below diagrams the basic patterns necessary to fully account for intergenerational savings and investment given a longer, healthier life. We believe that our attempt to capture this elusive notion of long-term savings, given improvements in the health of the previous generation, is one of the first such studies in this field. In order for us to emphasize this relationship, we pattern three different equations that would give us optimistic values of our independent variables of interest.

Methodology



We perform cross-sectional regressions to show the relationships between life expectancy, education, capital stock, and labor force. The pooled cross-sectional regressions performed at year 2000 provide an estimate of long-term elasticities. This is in contrast to the fixed-effects production function model that allowed for different intercepts for individual states and more, so represents short-term elasticities between our independent variables and output levels. As one can see from the data, there are more variations between states than within states, even given the substantial time period in our production function panel data (1970–2003). Over a significant time horizon, like our forty-year intergeneration analysis, one would expect to see this kind of larger variation growth within each state. For that reason we take the coefficients of our cross-sectional regressions as “long-term” elasticities.

This type of analysis has been problematic for many studies. For example, many health economists use variables like life expectancy to proxy for health status. Life expectancy in many developing countries is virtually time invariant and is only counted once every few decades. Therefore in fixed effects models, this variable and any other time invariant variables would be dropped. However, given the prosperity within the United States in this time period, we have unique and substantial state patterns that allow us to perform this in-depth study.

The cross-sectional analyses are shown below.

$$\log(Bachelor's) = \alpha + \beta \log(life\ expectancy\ at\ 65) + \theta \log(median\ earnings\ by\ degree) \quad \ldots\ldots \quad (4a)$$

Equation (4a) illustrates how improvements in health will feed back into educational investments. The median earnings by state data are taken from the census bureau.

Likewise, in order to relate how labor force and capital stock will be affected by long-term decisions to invest in education and health, we draw the following regressions.

$$\log(capital\ stock\ per\ employee) = \chi + \delta \log(Bachelor's) + \phi \log(life\ expectancy\ at\ 65) \quad \ldots\ldots \quad (4b)$$

and

$$\log(labor\ force\ participation\ rate) = \varphi + \gamma \log(life\ expectancy\ at\ 65) + \lambda \log\left(\frac{wages\ per\ employee}{median\ home\ price}\right) \quad \ldots\ldots \quad (4c)$$

where $\beta, \theta, \delta, \phi, \gamma, and\ \lambda$ are long-term elasticities to be measured. Wages and employment data are from BEA and BLS, respectively, and median home price is taken from the census bureau. All data is from year 2000. All other variables were from our production function data set and have been discussed in previous sections.

Our second step is to take these elasticities along with our optimistic life-expectancy-at-age-65 variable that we created to derive an optimistic forecast for bachelor's degree by state, followed by the computation of an optimistic forecast for labor force and capital stock. All coefficients of interest are displayed in the following table.

Reaction Functions

Variables		
Dependent	**Explanatory**	**Coefficient**
Log (Percentage of population with Bachelor Degree)	Log (Life expectancy at 65)	1.80** (3.95)
Log (Labor Force Participation Rate)	Log (Life expectancy at 65)	0.55** (2.87)
Log (Capital Stock per Employee)	Log (Percentage of population with Bachelor Degree)	0.56** (4.76)

**significant at the 1% level

Source: Milken Institute

The β coefficient, representing the long-run elasticity between life expectancy and education, is 1.8. We reason that the impact of this coefficient is time-varying, so we develop a time-varying pattern (2004–2050) that represents what we forecast the impact will be. This S-curve is a common technique for forecasting. From this computation, we are able to derive an optimistic forecast for our bachelor's degree variable by state. A similar pattern is derived for life expectancy's impact on labor-force participation rate (γ). This technique is not necessary for the capital stock reaction function (4b). Capital stock per employee already has the embedded S-curve because it utilizes the optimistic forecast of the bachelor's degree. The descriptions and computations of the variables are discussed in more detail in the following section.

The intergenerational analysis relies on the formation and application of the life expectancy series. For the optimistic forecast, we considered the construction of the baseline optimistic case and the careful analysis of the most recently available (1997–2003) six years of the NCHS complete life table data. Tying in our optimistic scenario of cancer disease incidence rates and heart disease prevalence rates from previous chapters, the two leading causes of death amongst chronic diseases, we form our expectation about the mortality rate of the over 65 population. We computed coefficients between mortality rates and life expectancy at 65, and used that coefficient, along with the historical trend over the past four decades, to determine that life expectancy in the year 2023 will increase by roughly 0.7 year when compared to the baseline; while by 2050, innovations to health and the focus on lifestyle will increase life expectancy at 65 by 1.7 years when compared again to the baseline.

Life expectancy feeds into decisions to invest in education. However, the impact will vary over time. We develop an S-curve that represents the magnitude of the impact of life expectancy at 65 on decisions to invest in bachelor's degrees and above, given the coefficient on our regression analysis. Generally, the greater impact of life expectancy should occur within the first twenty years and increase at a decreasing rate until 2050. We control for median earnings by educational attainment in the regression as well, since higher relative incomes will make acquiring higher educational degrees more appealing. Again, the optimistic life-expectancy-at-65 variable helps to generate our optimistic bachelor's degree variable.

Using our newly created optimistic bachelor's degree series, we take the next step and plug the series into a capital stock formation equation. Decisions to invest in capital stock (software, equipment, and structures) are determined by the percentage of population with higher education degrees. This will influence where they privately decide to invest monies. Although we had suspected that greater life expectancies would also increase investment, regression results were insignificant and dropped.

Like a bachelor's degree, life expectancy at 65 has a time-varying impact on the labor-force participation rate. We utilize the same pattern as before in the education reaction function. We simply alter the magnitude to coincide with the regression coefficient from the labor-force participation rate reaction function here. In our recursive model, decisions to invest in better health will have a positive and significant impact on a person's life, as well as workforce longevity. This model design departs from existing literature by not just projecting domestic regional markets but also relaying the spillover effects of health that have not been captured in any previous model. Better health enables a worker to remain in the labor pool longer. Feedback into the production function will demonstrate by how much this specific byway will increase each state's productivity.

The last variable necessary for our production function comparison is young dependents per capita. It was held constant; hence baseline and optimistic ratios derived from the Economy.com population projections are exactly the same.

Using the newly created optimistic series of life expectancy, education, labor force, and capital stock, along with the young dependents per capita data, we simply plug this back into the original production function with the state fixed effects. The gap between optimistic and baseline presents a difference of 17.59 percent by 2050. This translates to a pervasive underreporting of GDP by double-digit percentages by other models when they fail to account for the interaction of health to other variables.

The widening gap also serves as a way to compare the results from the intergenerational analysis with those in the nominal indirect and direct costs from previous chapters.

We have developed a modeling methodology that powerfully demonstrates health's contribution to economic growth. Our calculations simulate the impact of prevention, early detection and treatment of chronic disease for all fifty states. We believe that this analysis could help to change the paradigm of health, from a cost to an investment in promoting economic growth.